

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | | |
|---|--|--|---|
| (51) International Patent Classification ⁶ : C12N 15/53, 15/82, C12Q 1/26, 1/02, A01H 5/00 | | A1 | (11) International Publication Number: WO 97/49816 (43) International Publication Date: 31 December 1997 (31.12.97) |
| (21) International Application Number: PCT/US97/11295 (22) International Filing Date: 26 June 1997 (26.06.97) (30) Priority Data: 60/021,364 27 June 1996 (27.06.96) US (71) Applicant (for all designated States except US): E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): MAXWELL, Carl, Arthur [US/US]; 35 Mary Anita Court, Elkton, MD 21921 (US). SCOLNIK, Pablo, Ariel [US/US]; 120 Spotswood Lane, Kennett Square, PA 19348 (US). WITTENBACH, Vernon, Arie [US/US]; 609 Greenbank Road, Wilmington, DE 19808 (US). GUTTERIDGE, Steven [GB/US]; 4 Austin Road, Wilmington, DE 19810 (US). (74) Agent: FLOYD, Linda, A.; E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US). | | (81) Designated States: AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CU, CZ, EE, GE, HU, IL, IS, JP, KG, KP, KR, KZ, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, SL, TJ, TM, TR, TT, UA, US, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published With international search report. | |
| (54) Title: PLANT GENE FOR <i>P</i> -HYDROXYPHENYLPYRUVATE DIOXYGENASE | | | |
| <pre>1 CAAGAAACGNGTCCGCGACGGTCTGAGCGATGATCAGATCAAGGAGTGTGAGGAATTAGG. 61 GATTCTTTTAGACAGAGATGATCAAGGGAGCTTTCATCAAAATCTNCACAAAACCACTAGG 121 TGACAGCCCGAGGCTATTTATAGAGATAATCCAGAGCTTACGATCCATGATGAAGATGT 181 GGAAGGGAGGCGTTACCAAGATCGAGNATNTGTGCTTTCCGGAACGCAATT</pre> | | | |
| (57) Abstract <p>The invention relates to the isolation and modification of nucleic acid sequences encoding <i>p</i>-hydroxyphenylpyruvate dioxygenase enzyme from plants. These nucleic acid sequences were used to establish methods of identification of new herbicidal compounds that inhibit the activity of this enzyme, and to prepare new crop plants that are tolerant to the herbicidal action of inhibitors of this enzyme. Chimeric genes comprising nucleic acid fragments containing all or part of the nucleic acid sequences encoding <i>p</i>-hydroxyphenylpyruvate dioxygenase may be used to produce active plant <i>p</i>-hydroxyphenylpyruvate dioxygenase enzyme in microorganisms, and to cause the production of modified forms of the enzyme in plants that may render such plants tolerant to inhibitors of the enzyme.</p> | | | |

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|----|--------------------------|----|---------------------------------------|----|---|----|--------------------------|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | ML | Mali | TR | Turkey |
| BG | Bulgaria | HU | Hungary | MN | Mongolia | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MR | Mauritania | UA | Ukraine |
| BR | Brazil | IL | Israel | MW | Malawi | UG | Uganda |
| BY | Belarus | IS | Iceland | MX | Mexico | US | United States of America |
| CA | Canada | IT | Italy | NR | Niger | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NL | Netherlands | VN | Viet Nam |
| CG | Congo | KE | Kenya | NO | Norway | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NZ | New Zealand | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | PL | Poland | | |
| CM | Cameroon | KR | Republic of Korea | PT | Portugal | | |
| CN | China | KZ | Kazakhstan | RO | Romania | | |
| CU | Cuba | LC | Saint Lucia | RU | Russian Federation | | |
| CZ | Czech Republic | LI | Liechtenstein | SD | Sudan | | |
| DE | Germany | LK | Sri Lanka | SE | Sweden | | |
| DK | Denmark | LR | Liberia | SG | Singapore | | |
| EE | Estonia | | | | | | |

TITLEPLANT GENE FOR *p*-HYDROXYPHENYLPYRUVATE DIOXYGENASEFIELD OF THE INVENTION

This invention relates to the isolation and modification of nucleic acid
5 encoding *p*-hydroxyphenylpyruvate dioxygenase enzyme from plants. These
nucleic acid sequences were used to establish methods of identification of new
herbicidal compounds that inhibit the activity of this enzyme, and to prepare new
crop plants that are tolerant to the herbicidal action of inhibitors this enzyme.
Chimeric genes comprising nucleic acid fragments containing all or part of the
10 nucleic acid sequences encoding *p*-hydroxyphenylpyruvate dioxygenase may be
used to produce active plant *p*-hydroxyphenylpyruvate dioxygenase enzyme in
microorganisms, and to cause the production of modified forms of the enzyme in
plants that may render such plants tolerant to inhibitors of the enzyme.

BACKGROUND OF THE INVENTION

15 Bleaching herbicides affect plant chloroplasts by decreasing their
chlorophyll and carotenoid content. Several bleaching herbicides are known to
inhibit the enzyme phytoene desaturase, resulting in the accumulation of phytoene
in treated plants. However, compounds of the benzoyl cyclohexane-1,3-dione
type cause the accumulation of phytoene in plants but are not inhibitors of
20 phytoene desaturase *in vitro* (Sandmann, G., et al. (1990) *Pestic. Sci.* 30:353-355).
Subsequent work revealed that these compounds are effective inhibitors of
p-hydroxyphenylpyruvate dioxygenase (*p*-hydroxyphenylpyruvate: oxygen
oxidoreductase EC 1.13.11.27), a key enzyme in the biosynthesis of
plastoquinones and tocopherols (Schulz, A., et al. (1993) *FEBS Lett.*
25 318:162-166). Based on the observation that phytoene desaturase requires a
quinone as an electron acceptor, these authors postulated that by inhibiting
p-hydroxyphenylpyruvate dioxygenase, these herbicides act indirectly on
phytoene desaturase by blocking the biosynthesis of quinones.

The proposal that *p*-hydroxyphenylpyruvate dioxygenase is essential for
30 carotenoid biosynthesis has received support from genetic studies in the plant
model system *Arabidopsis thaliana*. Mutations in the *pds1* and *pds2* genetic loci
result in mutant plants that accumulate phytoene. However, genetic mapping of
these mutant genes indicates that they do not correspond to the gene encoding the
enzyme phytoene desaturase. The *pds1* mutation can be rescued by homogentisic
35 acid, the substrate of *p*-hydroxyphenylpyruvate dioxygenase. Therefore, this
mutation corresponds to a defect in the activity of *p*-hydroxyphenylpyruvate
dioxygenase (Norris, S. R., et al. (1995) *Plant Cell* 7:2139-2149).

In light of these disclosures, *p*-hydroxyphenylpyruvate dioxygenase is a promising new target for new herbicidal compounds. Research aimed at discovering new herbicides based on this mode of action would be greatly facilitated by the isolation of the plant gene encoding this enzyme and by the functional expression of this gene in transgenic organisms. For example, active enzyme produced in recombinant microorganisms could be used to establish screening methods for the identification of novel active compounds and to obtain structural and mechanistic information useful to guide further chemical synthesis. Furthermore, isolation of this gene would facilitate research aimed at generating mutant, herbicide-tolerant versions of the enzyme that may confer herbicide resistance to transgenic plants.

A partial sequence of an *Arabidopsis thaliana* cDNA with homology to corresponding mammalian sequences encoding *p*-hydroxyphenylpyruvate dioxygenase has been identified (GenBank Accession No. T20952), but this truncated sequence is insufficient to identify an active plant *p*-hydroxyphenylpyruvate dioxygenase. WO 96/38567 A2 addresses the utility that would be attached to a DNA sequence of a *p*-hydroxyphenylpyruvate dioxygenase gene, but there is no biochemical evidence of function associated with the sequences disclosed.

SUMMARY OF THE INVENTION

This invention pertains to the isolation and characterization of nucleic acid fragments encoding plant *p*-hydroxyphenylpyruvate dioxygenase enzymes. More specifically, this invention pertains to isolated nucleic acid fragments encoding the *p*-hydroxyphenylpyruvate dioxygenase enzymes from *Arabidopsis thaliana* and *Zea mays*.

This invention also pertains to the production of active plant *p*-hydroxyphenylpyruvate dioxygenase enzyme in *E. coli*. In one embodiment, a chimeric gene comprising a nucleic acid fragment encoding a polypeptide that possesses *p*-hydroxyphenylpyruvate dioxygenase activity, operably linked to regulatory sequences that direct gene expression in *E. coli*, is claimed. In another embodiment, a plasmid vector comprising said chimeric gene is disclosed. In yet another embodiment, a transformed *E. coli* comprising a chimeric gene consisting of a nucleic acid fragment encoding a polypeptide that possesses *p*-hydroxyphenylpyruvate dioxygenase activity is disclosed.

This invention also pertains to a method of identifying substances that inhibit the rate of the reaction of *p*-hydroxyphenylpyruvate dioxygenase enzyme. In one embodiment, the invention pertains to an assay for the detection of inhibitors of *p*-hydroxyphenylpyruvate dioxygenase wherein a polypeptide

derived from a transformed *E. coli* that displays *p*-hydroxyphenylpyruvate dioxygenase activity is incubated in the presence of a test substance. Following incubation, *p*-hydroxyphenylpyruvate dioxygenase enzymatic activity is measured wherein a reduction of enzymatic activity is indicative of the inhibitory capacity of the test substance. Enzymatic activity can be measured by any appropriate means, including but not limited to oxygen utilization, carbon dioxide release, homogentisate production, and loss of *p*-hydroxyphenylpyruvate. Results are quantified by radiometric, colorimetric or chromatographic means.

In another embodiment, this invention pertains to plants that are substantially tolerant to the application of at least one compound that inhibits the rate of the reaction of *p*-hydroxyphenylpyruvate dioxygenase. Plants may be rendered tolerant by overexpression of the wild-type *p*-hydroxyphenylpyruvate dioxygenase, by expression of a naturally-occurring resistant variant of this enzyme, or by expression of an altered form of *p*-hydroxyphenylpyruvate dioxygenase that is resistant to the action of compounds that are inhibitory to the wild-type enzyme.

A further embodiment of the invention is an isolated nucleic acid fragment comprising a member selected from the group consisting of:

- (a) an isolated nucleic acid fragment as set forth in SEQ ID NO:16;
- (b) an isolated nucleic acid fragment that is essentially similar to an isolated nucleic acid fragment as set forth in SEQ ID NO:16;
- and
- (c) an isolated nucleic acid fragment that is complementary to (a) or (b).

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the accompanying drawings and the sequence descriptions which form a part of this application.

Figure 1 presents a partial nucleic acid sequence of an expressed sequence tag (EST) bearing GenBank Accession No. T92052 obtained from an *Arabidopsis thaliana* cDNA library. This sequence was contained in clone 91B13T7 of the library.

Figure 2 presents the nucleic acid sequence of the cloned cDNA encoding a full-length form of *Arabidopsis thaliana* *p*-hydroxyphenylpyruvate dioxygenase enzyme, as it was initially determined (SEQ ID NO:2). Translation start and stop codons are underlined. Selected restriction sites are indicated.

Figure 3 presents the amino acid sequence comparison between full-length *p*-hydroxyphenylpyruvate dioxygenases from *Arabidopsis thaliana* (SEQ ID NO:15) and *Zea mays* (SEQ ID NO:11) and the *p*-hydroxyphenylpyruvate dioxygenase enzymes derived from human (SEQ ID NO:6, GenBank Acc. No. U29895), pig (SEQ ID NO:7, GenBank Acc. No. D13390), mouse (SEQ ID NO:8, GenBank Acc. No. D29987) and rat (SEQ ID NO:9, GenBank Acc. No. M18405). Asterisks indicate amino acid residues that are conserved across all six species. This figure was created using the Pileup program of GCG (Program Manual for the Wisconsin Package, Version 9.0-OpenVMS, December 1996, Genetics Computer Group, 575 Science Drive, Madison, WI, USA 53711).

Figure 4 is a diagram describing the construction of the intermediate plasmid vector pT7BlueR + PDO1.

Figure 5 is a diagram describing the construction of *E. coli* expression vector pE24CP1.

Applicants have provided a sequence listing in conformity with "Rules for the Standard Representation of Nucleotide and Amino Acid Sequences in Patent Applications" (Annexes I and II to the Decision of the President of the EPO, published in Supplement No. 2 to OJ EPO, 12/1992) and with 37 C.F.R. 1.821-1.825 and Appendices A and B ("Requirements for Application Disclosures Containing Nucleotides and/or Amino Acid Sequences").

SEQ ID NO:1 presents a partial nucleic acid sequence of an expressed sequence tag (EST) bearing GenBank Accession No. T92052 obtained from an *Arabidopsis thaliana* cDNA library. This sequence was contained in clone 91B13T7 of the library.

SEQ ID NO:2 presents the initial determination of the nucleic acid sequence and the deduced amino acid sequence of a cDNA encoding a full-length form of *Arabidopsis thaliana p*-hydroxyphenylpyruvate dioxygenase enzyme, as contained in plasmid pGBPPD2.

SEQ ID NO:3 presents the initially deduced amino acid sequence encoded by a cDNA for *Arabidopsis thaliana p*-hydroxyphenylpyruvate dioxygenase enzyme.

SEQ ID NOS:4 and 5 present the nucleotide sequences of a pair of complementary oligonucleotides (CAM 32 and CAM 33, respectively) used to facilitate subcloning and expression of the gene encoding *p*-hydroxyphenylpyruvate dioxygenase without the chloroplast transit sequence.

SEQ ID NO:6 presents the amino acid sequence of *p*-hydroxyphenylpyruvate dioxygenase enzyme derived from human (GenBank Acc. No. U29895).

SEQ ID NO:7 presents the amino acid sequence of *p*-hydroxyphenyl-pyruvate dioxygenase enzyme derived from pig (GenBank Acc. No. D13390).

SEQ ID NO:8 presents the amino acid sequence of *p*-hydroxyphenyl-pyruvate dioxygenase enzyme derived from mouse (GenBank Acc. No. D29987).

5 SEQ ID NO:9 presents the amino acid sequence of *p*-hydroxyphenyl-pyruvate dioxygenase enzyme derived from rat (GenBank Acc. No. M18405).

SEQ ID NO:10 presents the nucleic acid sequence and deduced amino acid sequence of the cloned cDNA encoding the *Zea mays p*-hydroxyphenylpyruvate dioxygenase enzyme, as contained in plasmid pMPDO.

10 SEQ ID NO:11 presents the deduced amino acid sequence of the cloned cDNA encoding the *Zea mays p*-hydroxyphenylpyruvate dioxygenase enzyme, as contained in plasmid pMPDO.

SEQ ID NO:12 presents the nucleic acid sequence and the deduced amino acid sequence of the truncated form of *Arabidopsis thaliana p*-hydroxyphenyl-pyruvate dioxygenase enzyme as contained in pE24CP1.

15 SEQ ID NO:13 presents the deduced amino acid sequence of the truncated form of *Arabidopsis thaliana p*-hydroxyphenylpyruvate dioxygenase enzyme as contained in pE24CP1.

SEQ ID NO:14 presents the revised nucleic acid sequence and the deduced amino acid sequence of the cloned cDNA encoding the full-length *Arabidopsis thaliana p*-hydroxyphenylpyruvate dioxygenase enzyme, as contained in plasmid pGBPPD2.

20 SEQ ID NO:15 presents the revised amino acid sequence deduced from the cDNA for the full length *Arabidopsis thaliana p*-hydroxyphenylpyruvate dioxygenase enzyme.

25 SEQ ID NO:16 presents the nucleic acid sequence determined from a portion of a cDNA from *Vernonia galamensis*, as contained in clone vs1.pk0015.b2.

DETAILS OF THE INVENTION

BIOLOGICAL DEPOSITS

30 The following biological materials have been deposited under the terms of the Budapest Treaty at American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, MD 20852, and bear the following accession numbers:

35

| <u>Depositor Identification</u> | | <u>Int'l. Depository</u> | |
|---------------------------------|----------------|--------------------------|------------------------|
| <u>Host Strain</u> | <u>Plasmid</u> | <u>Accession Number</u> | <u>Date of Deposit</u> |
| <i>E. coli</i> BL21(DE3) | pE24CP1 | ATCC 98083 | June 25, 1996 |
| N/A | pGBPPD2 | ATCC 97622 | June 25, 1996 |
| N/A | pMPDO | ATCC 209120 | June 12, 1997 |

Definitions

In the context of this disclosure, a number of terms shall be utilized. As used herein, the term "nucleic acid" refers to a large molecule which can be single-stranded or double-stranded, composed of monomers (nucleotides) containing a sugar, phosphate and either a purine or pyrimidine. A "nucleic acid fragment" is a portion of a given nucleic acid molecule. As used herein, "DNA" (deoxyribonucleic acid) is the genetic material, whereas "RNA" (ribonucleic acid) is involved in the transfer of the information encoded by the DNA into proteins and polypeptides. A "genome" is the entire body of genetic material contained in each cell of an organism. The term "nucleotide sequence" refers to a polymer of DNA or RNA which can be single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases capable of incorporation into DNA or RNA polymers.

As used herein, "essentially similar" refers to DNA sequences that may involve base changes that do not cause a change in the encoded amino acid or which involve base changes which may alter one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence. It is therefore understood that the invention encompasses more than the specific exemplary sequences. Modifications to the sequence, such as deletions, insertions, or substitutions in the sequence which produce "silent changes" (i.e., those that do not substantially affect the functional properties of the resulting protein molecule) are also contemplated. For example, alteration(s) in the gene sequence which reflects the degeneracy of the genetic code, or which result in the production of a chemically equivalent amino acid at a given site, are contemplated; thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a biologically equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the protein molecule would also not be

expected to alter the activity of the protein. In some cases, it may in fact be desirable to make mutants of the sequence in order to study the effect of alteration on the biological activity of the protein. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Moreover, the skilled artisan recognizes that "essentially similar" sequences encompassed by this invention are also defined by their ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65°C), with the sequences exemplified herein.

"Gene" refers to a nucleic acid fragment that encodes a specific protein, including regulatory sequences preceding (5' non-coding) and following (3' non-coding) the coding region. "Native" gene refers to the gene as found in nature with its own regulatory sequences. "Chimeric" gene refers to a gene comprising heterogeneous regulatory and coding sequences. "Endogenous" gene refers to the native gene normally found in its natural location in the genome. A "foreign" gene refers to a gene not normally found in the host organism but that is introduced by gene transfer.

"Coding sequence" refers to a DNA sequence that codes for a specific protein and excludes the non-coding sequences.

"Initiation codon" and "termination codon" refer to a unit of three adjacent nucleotides in a coding sequence that specifies initiation and termination, respectively, of protein synthesis (mRNA translation). "Open reading frame" refers to the amino acid sequence encoded between translation initiation and termination codons of a coding sequence.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript. "Messenger RNA" (mRNA) refers to RNA that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA, one strand of which is complementary to and derived from mRNA by reverse transcription. "Sense RNA" refers to RNA transcript that includes the mRNA.

As used herein, "regulatory sequences" are nucleotide sequences that control the transcription or expression of a coding sequence located upstream (5'), within, or downstream (3') to the coding sequence, act in conjunction with the protein biosynthetic apparatus of the cell and include promoters, translation leader sequences, transcription termination sequences, and polyadenylation sequences.

“Promoter” refers to a DNA sequence in a gene, usually upstream (5') to its coding sequence, which controls the expression of the coding sequence by providing the recognition for RNA polymerase and other factors required for proper transcription. A promoter may also contain DNA sequences that are involved in the binding of protein factors which control the effectiveness of transcription initiation in response to physiological or developmental conditions. In the case of eukaryotic organisms, it may also contain enhancer elements.

An “enhancer element” is a DNA sequence which can stimulate promoter activity. It may be an innate element of the promoter or a heterologous element inserted to enhance the activity level and tissue-specificity of a promoter.

“Constitutive promoters” refer to those enhancer elements that direct gene expression in all tissues and at all times. “Organ-specific” or “development-specific” promoters as referred to herein are those that direct gene expression almost exclusively in specific organs, such as leaves or seeds, or at specific development stages in an organ, such as in early or late embryogenesis, respectively.

The term “operably linked” refers to nucleic acid sequences on a single nucleic acid molecule which are associated so that the function of one is affected by the other. For example, a promoter is operably linked with a structural gene (i.e., a gene encoding *p*-hydroxyphenylpyruvate dioxygenase, as disclosed herein) when it is capable of affecting the expression of that structural gene (i.e., that the structural gene is under the transcriptional control of the promoter).

The term “expression”, as used herein, is intended to mean the production of the protein product encoded by a gene. More particularly, “expression” refers to the transcription and stable accumulation of the sense RNA (mRNA) derived from the nucleic acid fragment(s) of the invention that, in conjunction with the protein apparatus of the cell, results in altered levels of protein product.

“Overexpression” refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. “Altered levels” refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms. “Facilitating expression” refers to steps and conditions for culturing host cells containing the desirable gene to yield an increased production of the enzyme. For example, addition of a chemical inducer specific to the particular promoter operably linked to the gene facilitates expression of the encoded enzyme. This is measured relative to the production levels of an untreated gene.

The "3' non-coding sequences" refers to the DNA sequence portion of a gene that contains a polyadenylation signal and any other regulatory signal capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor.

The "translation leader sequence" refers to that DNA sequence portion of a gene between the promoter and coding sequence that is transcribed into RNA and is present in the fully processed mRNA upstream (5') of the translation start codon. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability, or translation efficiency.

"Transformation" herein refers to the transfer of a foreign gene into the genome of a host organism and its genetically stable inheritance. Bacterial transformation can proceed by any of several methods well known in the art, including calcium chloride-mediated transformation and electroporation. Examples of methods of plant transformation include *Agrobacterium*-mediated transformation and particle-accelerated or "gene gun" transformation technology (U.S. Patent No. 4,945,050).

"Host cell" refers to the cell that is transformed with the introduced genetic material.

"Plasmid vector" refers to a double-stranded, closed circular, extra-chromosomal DNA molecule.

"Tolerant" or "tolerance" refers to a condition whereby a cell or an organism is able to withstand the effect of application of a compound or composition at a concentration or application rate that causes a demonstrable effect in or against cells or organisms that are not tolerant. For example, the growth or survival of a plant that is tolerant to application of a herbicidal compound or composition will be less affected than the growth or survival of a plant that is not tolerant to application of the herbicidal compound or composition.

Cloning of Plant Genes Encoding *p*-Hydroxyphenylpyruvate Dioxygenase

The *p*-hydroxyphenylpyruvate dioxygenases from plants are a promising new class of targets for new herbicidal compounds. In order to be able to study this enzyme in detail, and to have available supplies of enzyme for inhibitor screening, cDNA clones encoding plant *p*-hydroxyphenylpyruvate dioxygenases were identified. These nucleic acid fragments are useful for the production of their encoded enzymes, for isolation of clones from additional plant sources that encode other *p*-hydroxyphenylpyruvate dioxygenase enzymes, and for understanding the biochemical and structural properties of these enzymes.

Nucleic acid fragments comprising nucleotide sequences that encode different forms of the enzyme *p*-hydroxyphenylpyruvate dioxygenase from the plant *Arabidopsis thaliana* have now been isolated. Subsequently, these nucleotide sequences were expressed in *E. coli* cells and shown to direct the synthesis of plant *p*-hydroxyphenylpyruvate dioxygenase enzymes.

An automated search of nucleotide sequences contained in a database representing an *Arabidopsis* cDNA library for sequences homologous to other known, non-plant *p*-hydroxyphenylpyruvate dioxygenase genes revealed the plasmid cDNA clone 91B13T7. This cDNA was obtained from the Arabidopsis Seed Stock Center at Ohio State University. Plasmid DNA suitable for nucleotide sequence determination was prepared and the nucleotide sequence of the plasmid insert was determined. The resulting sequence was not interpretable, suggesting possible contamination of the plasmid sample by an extraneous nucleic acid. This assumption was confirmed by digesting the plasmid DNA sample with restriction enzymes and separating the resulting nucleic acid fragments by agarose gel electrophoresis. This analysis revealed the presence of nucleic acid fragments that could not be derived from the plasmid carrying the putative *p*-hydroxyphenylpyruvate dioxygenase fragment. Furthermore, a search of the publically available nucleic acid sequence databases revealed that the *Arabidopsis thaliana* sequence reported for cDNA clone 91B13T7 corresponded to a truncated cDNA (Figure 1). Based on publically available mammalian cDNA sequence information for *p*-hydroxyphenylpyruvate dioxygenase, the minimum length expected for a cDNA encoding a complete *p*-hydroxyphenylpyruvate dioxygenase enzyme is 1 kb (Table 1).

Table 1
Predicted cDNA Length for Sequences
Encoding *p*-Hydroxyphenylpyruvate Dioxygenase

| Organism | Amino Acid Residues | Minimum cDNA (kb) |
|------------------------|---------------------|-------------------|
| Human | 392 | 1.176 |
| Pig | 392 | 1.176 |
| <i>Pseudomonas</i> sp. | 357 | 1.071 |

Therefore, based on the expected length of a cDNA capable of encoding a functional *p*-hydroxyphenylpyruvate dioxygenase, the *Arabidopsis thaliana* sequence obtained from the public database was insufficient to encode a full-length, active *p*-hydroxyphenylpyruvate dioxygenase enzyme. Therefore, a cDNA with the capacity to encode a full-length enzyme *Arabidopsis thaliana* was cloned,

as described herein. A 400 bp segment of the insert of plasmid 91B13T7 was liberated by digestion with restriction enzymes and used to screen a cDNA library prepared from norflurazon-treated *Arabidopsis thaliana* seedlings (Scolnik, P. A., and Bartley, G. E. (1994) *Plant Physiol.* 104:1469-1470). Several clones showing positive hybridization to this probe were sequenced. The initial determination of the sequence of the longest cDNA clone obtained from this effort is shown in Figure 2 and in SEQ ID NO:2. During the course of subsequent work with this clone it became necessary to confirm certain features of the sequence. A corrected sequence of this cDNA is presented in SEQ ID NO:12.

The sequence reported in Figure 2 indicates that this cDNA has the capacity to encode a protein of MW 48,841 which, as shown in Figure 3, has a high level of homology to *p*-hydroxyphenylpyruvate dioxygenase enzymes from other eukaryotes.

A cDNA capable of encoding a full-length *p*-hydroxyphenylpyruvate dioxygenase has also been obtained from corn. This cDNA, contained in plasmid pMPDO, was identified in a corn cDNA library using an approximately 900 base pairs portion of the *Arabidopsis* cDNA as a probe. The predicted amino acid sequence that is encoded by the corn cDNA is also compared to *p*-hydroxyphenylpyruvate dioxygenase enzymes from other eukaryotes in Figure 3.

A cDNA library was prepared from messenger RNA isolated from developing seeds of *Vernonia galamensis*. Random sequencing of the clones contained in the library identified a probable clone, designated vs1.pk0015.b2, for the *p*-hydroxyphenylpyruvate dioxygenase from this plant. The 513 bp expressed sequence tag (EST) is presented in SEQ ID NO:16.

Expression of the *Arabidopsis thaliana* cDNA Encoding *p*-Hydroxyphenylpyruvate Dioxygenase in *E. coli*

The nucleic acid fragments of the instant invention encoding a plant *p*-hydroxyphenylpyruvate dioxygenase enzymes can be operably linked to suitable regulatory sequences, thereby creating chimeric genes that can be used to direct expression of the enzyme in transgenic organisms. These transgenic organisms include, but are not limited to: plants (*Plant Molecular Biology*; Croy, R. R. D., Ed.; Bios Scientific Publishers; 1993); microorganisms, including *Escherichia coli* (Gold, L. (1990) *Methods in Enzymology* 185:11), *Bacillus subtilis* (Henner; D. J. (1990) *Methods in Enzymology* 185:199), yeast (Gellissen, G., et al. (1992) *Antonie Leeuwenhoek* 62:79), and fungi, including members of the genus *Aspergillus* (Devchand, M. and Gwynne, D. I. (1991) *J. Biotechnol.* 17:3); and insect cells containing recombinant baculoviruses (Lukow, V. A. and Summers, M. D. (1988) *Bio/Technology* 6:47).

One skilled in the art can isolate the coding sequences from the fragments of the invention by using or creating sites for restriction endonucleases, as described in Sambrook, J., et al. ((1989) *Molecular Cloning, A Laboratory Manual*, 2nd ed.; Cold Spring Harbor Laboratory Press; hereinafter "Maniatis"). Alternatively, polymerase chain reaction (PCR) techniques can be employed to isolate and/or modify the fragments of the invention (Newton, C. R. and Graham, A. (1994) *PCR*; Bios Scientific Publishers).

Arabidopsis p-hydroxyphenylpyruvate dioxygenase was expressed in *E. coli* under control of a T7 promoter in a strain expressing T7 RNA polymerase (Studier, F. W., et al. (1990) *Methods in Enzymology* 185:60). Promoters other than T7 are commonly used in expression vectors and could be substituted for protein expression in *E. coli*. Examples of alternative promoters include, but are not limited to, *trp* (Yansura, D. G. and Henner, D. J. (1990) *Methods in Enzymology* 185:54), P_L (Remaut, E. et al. (1981) *Gene* 15:81), *tac* (Amann, E. et al. (1983) *Gene* 25:167), *trc* (Amann, E. et al. (1988) *Gene* 69:301), and promoters such as *lacUV5*, *lpp*, P_R , and hybrid and tandem promoters constructed to combine specific features to increase strength or regulation capacity (Balbas, P. and Bolivar, F. (1990) *Methods in Enzymology* 185:14).

Biochemical Evidence of Enzymatic Function

The enzyme *p*-hydroxyphenylpyruvate dioxygenase catalyzes the reaction of *p*-hydroxyphenylpyruvate with molecular oxygen to give homogentisate and CO₂. The enzyme can be assayed by measuring oxygen utilization (Hager, S. E., et al. (1957) *J. Biol. Chem.* 225:935-947), CO₂ release or homogentisate production from radioactive labeled *p*-hydroxyphenylpyruvate (Lindblad, B. (1971) *Clin. Chem. Acta* 34:113-121), loss of the *p*-hydroxyphenylpyruvate (Lin, E. C. C. et al. (1958) *J. Biol. Chem.* 233:668-673), or formation of homogentisate using a colorimetric assay (Fellman, J. H. et al. (1972) *Biochim. Biophys. Acta* 284:90-100) or UV detection following HPLC or a similar chromatographic separation technique. The activity of *p*-hydroxyphenylpyruvate dioxygenase may also be measured in a coupled assay in which the initial product, homogentisate, is oxidized by homogentisate dioxygenase; formation of maleylacetoacetate determined by measuring absorbance at 330 nm (Fernández-Cañón, J. M. and Peñalva, M. A. (1997) *Anal. Biochem.* 245:218-221).

An alternative to any of the kinetic assays for *p*-hydroxyphenylpyruvate dioxygenase is an end-point or fixed-time assay. The procedure is based on the conversion of unconverted substrate, *p*-hydroxyphenylpyruvate to its enediol tautomer by tautomerase in the presence of borate ions and measurement of the characteristic 308 nm peak of the tautomer (Lin, E. C. C. et al. (1958) *J. Biol.*

Chem. 233:668-673). The procedure involves the addition of enough *p*-hydroxyphenylpyruvate dioxygenase to consume ~80% of the organic substrate over 1 hour in 200 μ L of assay buffer, which in this case is a 50 mM Tris, pH 7.4, 0.10 mM *p*-hydroxyphenylpyruvic acid, 1.75 mM ascorbate and 1.25 mM EDTA. After 1 hr the reaction is quenched by the addition of 100 μ L of 0.8 M borate, pH 7.3, containing 1000 ppb of a *p*-hydroxyphenylpyruvate dioxygenase inhibitor and 0.25 μ L of 6.1 mg/mL of tautomerase. The absorbance at 308 nm is read after a 30 min incubation and is stable thereafter for 2 hr. The advantage of this assay over the kinetic procedure is that the *p*-hydroxyphenylpyruvate dioxygenase is not required to oxidize the substrate in the presence of high concentrations of borate, a condition that might interfere with the mode of action of inhibitors. Furthermore the assay produces essentially a stable binary indication of *p*-hydroxyphenylpyruvate dioxygenase inhibition, and is well-suited for applications which require a high-throughput of samples and assays.

The enzyme encoded by the nucleic acid fragments and overexpressed in *E. coli* can be extracted in any conventional buffer used for extracting soluble plant enzymes. Although a large amount of an overexpressed protein is often insoluble, the amount that is soluble represents can represent as much as 50% of the total soluble protein. Soluble overexpressed protein has high *p*-hydroxyphenylpyruvate dioxygenase activity and is easily extracted. Likewise, it may be possible to resolubilize an insoluble overexpressed protein in an active form under appropriate conditions, since addition of sarkosyl (sodium N-lauroylsarcosinate) to the extraction buffer appeared to increase the amount of the overexpressed protein extracted. For optimum activity, a reducing agent such as ascorbate or reduced glutathione should be present as well as a source a ferrous ion.

An overexpressed enzyme can be assayed using all the techniques described above for measuring *p*-hydroxyphenylpyruvate dioxygenase activity, while only the techniques using labeled *p*-hydroxyphenylpyruvate can be used to measure activity in crude plant extracts. Therefore, the availability of an overexpressed enzyme greatly facilitates the development of high capacity screens to identify inhibitors of the enzyme. Potential inhibitors are evaluated for their capacity to reduce the rate of the reaction of the enzyme, resulting in reduced oxygen uptake and CO₂ release, and lower rates of formation of homogentisate and loss of *p*-hydroxyphenylpyruvate. Applicants have demonstrated that at least one of the instant nucleic acid fragments can be overexpressed in *E. coli* cells, resulting in production of a protein that catalyzes the conversion of *p*-hydroxyphenylpyruvate to homogentisate with the release of CO₂. Furthermore, it has been shown that this activity is inhibited by commercial herbicides known to

inhibit *p*-hydroxyphenylpyruvate dioxygenase. Finally, an overexpressed enzyme can be used in a high capacity assay to identify compounds that inhibit the enzymatic activity of *p*-hydroxyphenylpyruvate dioxygenase. Such compounds may serve as herbicides.

5 Preparation of Plants Tolerant to Inhibitors of *p*-Hydroxyphenylpyruvate Dioxygenase

This invention embodies plants which are resistant or at least tolerant to herbicides that target the *p*-hydroxyphenylpyruvate dioxygenase enzyme at levels which are normally inhibitory to the naturally occurring *p*-hydroxyphenylpyruvate
10 dioxygenase enzyme. This altered *p*-hydroxyphenylpyruvate dioxygenase activity is conferred by (1) overexpression of the wild-type *p*-hydroxyphenylpyruvate dioxygenase enzyme, or (2) expression of a DNA molecule encoding a herbicide-tolerant enzyme. The said enzyme may be a modified form of an *p*-hydroxy-phenylpyruvate dioxygenase enzyme that occurs naturally in a eukaryote or
15 prokaryote, or a modified form of an *p*-hydroxyphenylpyruvate dioxygenase enzyme that naturally occurs in a plant, or a herbicide tolerant enzyme that naturally occurs in a prokaryote (Duke et al. *Herbicide Resistant Crops*; Lewis: Boca Raton; 1994). An effective amount of gene expression to render the cells of the plant tissue substantially tolerant to the herbicide depends on whether the gene
20 codes for an unaltered *p*-hydroxyphenylpyruvate dioxygenase gene or a mutant or altered form of the gene that is less sensitive to the herbicides. Expression of an unaltered plant *p*-hydroxyphenylpyruvate dioxygenase gene in an effective amount is that amount that provides for a 2- to 10-fold increase in herbicide tolerance. Plants encompassed by the invention include monocotyledoneous and
25 dicotyledoneous plants. Preferred are those plants which would be potential targets for *p*-hydroxyphenylpyruvate dioxygenase-inhibiting herbicides, particularly agronomically important crops such as maize and other cereal crops.

Increased levels of expression of *p*-hydroxyphenylpyruvate dioxygenase activity, from two to ten or more times the natively expressed amount, would be
30 sufficient to overcome growth inhibition caused by the herbicide. Plants containing such altered *p*-hydroxyphenylpyruvate dioxygenase enzyme activity can be obtained by direct selection in plants. This method is known in the art. See, e.g., U.S. Patent No. 5,162,602. U.S. Patent No. 4,761,373, and references cited therein.

35 Overexpression of *p*-hydroxyphenylpyruvate dioxygenase also can be accomplished by stably transforming a host plant cell with a chimeric DNA molecule comprising a promoter capable of driving expression of an associated coding sequence in a plant cell and operably linked to a homologous or

heterologous coding sequence encoding *p*-hydroxyphenylpyruvate dioxygenase. A "homologous" *p*-hydroxyphenylpyruvate dioxygenase gene is isolated from an organism taxonomically identical to the target plant cell, whereas a "heterologous" *p*-hydroxyphenylpyruvate dioxygenase gene is obtained from an organism taxonomically distinct from the target plant.

The expression of foreign genes in plants is well-established (De Blaere et al., (1987) *Meth. Enzymol.* 143:277-291). Promoters utilized to drive gene expression in transgenic plants or plant cells (i.e., those capable of driving expression of the associated coding sequences such as *p*-hydroxyphenylpyruvate dioxygenase in plant cells, include those directing the 19S and 35S transcripts in Cauliflower mosaic virus (Odell et al., (1985) *Nature* 313:810-812; Hull et al., (1987) *Virology* 86:482-493), small subunit of ribulose 1,5-bisphosphate carboxylase (Morelli et al., (1985) *Nature* 315:200-204; Broglie et al., (1984) *Science* 224:838-843; Herrera-Estrella et al., (1984) *Nature* 310:115-120; Coruzzi et al., (1984) *EMBO J.* 3:1671-1679; Faciotti et al., (1985) *Bio/Technology* 3:241 and chlorophyll *a/b* binding protein (Lamppa et al., (1986) *Nature* 316:750-752); nopaline synthase promoters (Depicker et al. (1982) *J. Mol. App. Genet.* 1:561-573; An et al. (1990) *Plant Cell* 2:225-233). The chimeric DNA construct(s) of the invention may contain multiple copies of a promoter or multiple copies of the *p*-hydroxyphenylpyruvate dioxygenase coding sequences. In addition, the construct(s) may include coding sequences for selectable markers and coding sequences for other peptides such as signal or transit peptides. The preparation of such constructs is within the ordinary level of skill in the art. Resistance to inhibitors of the plant carotenoid biosynthesis pathway, which is also targeted by *p*-hydroxyphenylpyruvate dioxygenase inhibitors, has been achieved by expressing a bacterial gene encoding phytoene desaturase driven by the CaMV promoter (Misawa et al., (1994) *Plant. J.* 4:481-490).

Transit peptides may be fused to the *p*-hydroxyphenylpyruvate dioxygenase coding sequence in the chimeric DNA constructs of the invention to direct transport of the expressed *p*-hydroxyphenylpyruvate dioxygenase enzyme to the desired site of action. Examples of transit peptides include the chloroplast transit peptides such as those described in Von Heijne et al., (1991) *Plant Mol. Biol. Rep.* 9:104-126; Mazur et al., (1987) *Plant Physiol.* 85:1110; Vorst et al., (1988) *Gene* 65:59; and mitochondrial transit peptides such as those described in Boutry et al., (1987) *Nature* 328:340-342.

It is envisioned that the introduction of enhancers or enhancer-like elements into other promoter constructs will also provide increased levels of primary transcription to accomplish the invention. These would include viral enhancers

such as that found in the 35S promoter (Odell et al., (1988) *Plant Mol. Biol.* 10:263-272), enhancers from the opine genes (Fromm et al., (1989) *Plant Cell* 1:977-984), or enhancers from any other source that result in increased transcription when placed into a promoter operably linked to the nucleic acid
5 fragment of the invention.

Introns isolated from the maize Adh-1 and Bz-1 genes (Callis et al., (1987) *Genes Dev.* 1:1183-1200), and intron 1 and exon 1 of the maize Shrunken-1 (sh-1) gene (Maas et al., (1991) *Plant Mol. Biol.* 16:199-207) may also be of use to increase expression of introduced genes. Results with the first intron of the maize
10 alcohol dehydrogenase (Adh-1) gene indicate that when this DNA element is placed within the transcriptional unit of a heterologous gene, mRNA levels can be increased by 6.7-fold over normal levels. Similar levels of intron enhancement have been observed using intron 3 of a maize actin gene (Luehrsen, K. R. and Walbot, V., (1991) *Mol. Gen. Genet.* 225:81-93). Enhancement of gene
15 expression by Adh1 intron 6 (Oard et al., (1989) *Plant Cell Rep* 8:156-160) has also been noted. Exon 1 and intron 1 of the maize sh-1 gene have been shown to individually increase expression of reporter genes in maize suspension cultures by 10 and 100-fold, respectively. When used in combination, these elements have been shown to produce up to 1000-fold stimulation of reporter gene expression
20 (Maas et al., (1991) *Plant Mol. Biol.* 16:199-207).

Any 3' non-coding region capable of providing a polyadenylation signal and other regulatory sequences that may be required for proper expression can be used to accomplish the invention. This would include the 3' end from any storage
25 protein such as the 3' end of the 10kd, 15kd, 27kd and alpha zein genes, the 3' end of the bean phaseolin gene, the 3' end of the soybean β -conglycinin gene, the 3' end from viral genes such as the 3' end of the 35S or the 19S cauliflower mosaic virus transcripts, the 3' end from the opine synthesis genes, the 3' ends of ribulose 1,5-bisphosphate carboxylase or chlorophyll a/b binding protein, or 3' end sequences from any source such that the sequence employed provides the
30 necessary regulatory information within its nucleic acid sequence to result in the proper expression of the promoter/coding region combination to which it is operably linked. There are numerous examples in the art that teach the usefulness of different 3' non-coding regions (for example, see Ingelbrecht et al., (1989) *Plant Cell* 1:671-680).

35 Various methods of introducing a DNA sequence (i.e., of transforming) into eukaryotic cells of higher plants are available to those skilled in the art (see EPO publications 0 295 959 A2 and 0 138 341 A1). Such methods include high-velocity ballistic bombardment with metal particles coated with the nucleic acid

constructs (see Klein et al., (1987) *Nature* (London) 327:70-73. and see U.S. Patent No. 4,945,050), as well as those based on transformation vectors based on the Ti and Ri plasmids of *Agrobacterium* spp., particularly the binary type of these vectors. Ti-derived vectors transform a wide variety of higher plants, including
5 monocotyledonous and dicotyledonous plants, such as soybean, cotton and rape seed (Pacciotti et al., (1985) *Bio/Technology* 3:241; Byrne et al., (1987) *Plant Cell, Tissue and Organ Culture* 8:3; Sukhapinda et al., (1987) *Plant Mol. Biol.* 8:209-216; Lorz et al., (1985) *Mol. Gen. Genet.* 199:178-182; Potrykus et al., (1985) *Mol. Gen. Genet.* 199:183-188).

10 Other transformation methods are available to those skilled in the art, such as direct uptake of foreign DNA constructs (see EPO publication 0 295 959 A2), and techniques of electroporation (see Fromm et al., (1986) *Nature* (London) 319:791-793). Once transformed, the cells can be regenerated by those skilled in the art. Also relevant are several recently described methods of introducing
15 nucleic acid fragments into commercially important crops, such as rapeseed (see De Block et al., (1989) *Plant Physiol.* 91:694-701), sunflower (Everett et al., (1987) *Bio/Technology* 5:1201-1204), soybean (McCabe et al., (1988) *Bio/Technology* 6:923-926; Hinchey et al., (1988) *Bio/Technology* 6:915-922; Chee et al., (1989) *Plant Physiol.* 91:1212-1218; Christou et al., (1989) *Proc. Natl. Acad. Sci USA* 86:7500-7504; EPO Publication 0 301 749 A2), and corn
20 (Gordon-Kamm et al., (1990) *Plant Cell* 2:603-618; and Fromm et al., (1990) *Bio/Technology* 8:833-839).

Altered *p*-hydroxyphenylpyruvate dioxygenase enzyme activity may also be achieved through the generation or identification of modified forms of the isolated
25 eukaryotic *p*-hydroxyphenylpyruvate dioxygenase coding sequence having at least one amino acid substitution, addition or deletion which encodes an altered *p*-hydroxyphenylpyruvate dioxygenase enzyme resistant to a herbicide that inhibits the unaltered, naturally occurring form. Genes encoding such enzymes can be obtained by numerous strategies known in the art. A first general strategy
30 involves direct or indirect mutagenesis procedures on microbes (e.g., *E. coli*, *S. cerevisiae* (Miller, (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; Davis et al., (1980) *Advanced Bacterial Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; Sherman et al., (1983) *Methods in Yeast Genetics*, Cold Spring Harbor
35 Laboratory, Cold Spring Harbor NY; and U.S. Patent No. 4,975,374) and cyanobacteria (Bryant, *The Molecular Biology of Cyanobacteria*; Kluwer Academic Publishers: Boston, 1995). A second method of obtaining mutant herbicide-resistant alleles of the eukaryotic *p*-hydroxyphenylpyruvate dioxygenase

enzyme involves direct selection in plants. For example, the effect of inhibitors on the growth of plants such as *Arabidopsis*, soybean, or maize may be determined by plating seeds sterilized by art-recognized methods on plates on a simple minimal salts medium containing increasing concentrations of the inhibitor. The lowest dose at which significant growth inhibition can be reproducibly detected is used for subsequent experiments. Mutagenesis of plant material may be utilized to increase the frequency at which resistant alleles occur in the selected population. Mutagenized seed material can be derived from a variety of sources, including chemical or physical mutagenesis of seeds, or chemical or physical mutagenesis of pollen (Neuffer, In *Maize for Biological Research*, Sheridan, ed. Univ. Press, Grand Forks, ND., pp. 61-64 (1982)), which is then used to fertilize plants and the resulting M1 mutant seeds collected. Typically, for *Arabidopsis*, M2 seeds (i.e., progeny seeds of plants grown from seeds mutagenized with chemicals, such as ethyl methane sulfonate, or with physical agents, such as gamma rays or fast neutrons) are plated at densities of up to 10,000 seeds/plate (10 cm diameter) on minimal salts medium containing an appropriate concentration of inhibitor. Seedlings that continue to grow and remain green 7-21 days after plating are transplanted to soil and grown to maturity and seed set. Progeny of these seeds are tested for resistance to the herbicide. If the resistance trait is dominant, plants whose seed segregate 3:1 (resistant:sensitive) are presumed to have been heterozygous for the resistance at the M2 generation. Plants that give rise to all resistant seed are presumed to have been homozygous for the resistance at the M2 generation. Such mutagenesis on intact seeds and screening of their M2 progeny seed can also be carried out on other species, for instance soybean (see, e.g., U.S. Patent No. 5,084,082). Mutant seeds to be screened for herbicide tolerance can also be obtained as a result of fertilization with pollen mutagenized by chemical or physical means.

EXAMPLE 1

Cloning of a cDNA for *Arabidopsis thaliana*

p-Hydroxyphenylpyruvate Dioxygenase

The plasmid containing the *Arabidopsis thaliana* 91B13T7 expressed sequence tag (Newman et al., (1994) *Plant Physiol* 106:1241-1255) was digested with the restriction enzymes *Bam*HI and *Eco*RI, and the resulting 400 bp fragment was used to screen a lambda phage cDNA library of *Arabidopsis thaliana* seedlings (Scolnik, P. A. and Bartley, G. E. (1994) *Plant Physiol.* 104:1469-1470) according to the following protocol.

E. coli KW251 cells were grown overnight in Luria Broth ("LB") containing 0.2% maltose and 10 mM MgSO₄. Cells were pelleted by centrifugation and

resuspended in 10 mM MgSO₄ to an OD₆₀₀ of 0.5. Cell aliquots (0.8 mL) were mixed with 0.1 mL of diluted phage samples and 7 mL of top agarose (0.7% agarose in LB containing 10 mM MgSO₄) at 45°C, and plated onto 150 mm Petri dishes containing LB agar. Phage plaques became visible in 5-7 h, at which point the plates were placed at 4°C.

Phage plaques were transferred to nitrocellulose filters according to standard techniques, and the filters were hybridized to ³²P-radiolabeled probe prepared according to the method of Feinberg and Vogelstein ((1983) *Anal. Biochem.* 132:6-13), using the hybridization conditions of Berlyn et al.((1989) *Proc. Natl. Acad. Sci.* 86:4604-4608). After exposure to X-ray film for 48 h, 12 positive plaques were eluted, plated, and hybridized under the same conditions. A total of 9 plaques that retained positive signals in this second round of hybridization were subjected to *in vivo* excision using the Exassist/SOLR™ system according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). DNA from the plasmids resulting from *in vivo* excision of positive plaques was prepared for DNA sequencing using the Wizard Plus™ kit (Promega, Madison, WI). Eight of the clones that were sequenced showed strong conservation with available *p*-hydroxyphenylpyruvate dioxygenase sequences, whereas the remaining clone did not correspond to a *p*-hydroxyphenylpyruvate dioxygenase. Alignment with known *p*-hydroxyphenylpyruvate dioxygenase sequences also revealed that two of the clones correspond to 0.3 kbp fragments from the 3' end of the transcript, and another two to 1.2 kbp fragments from the 5' end of the transcript. One clone of each was used to assemble a 1.5 kbp cDNA by ligating at the internal *Nhe*I restriction site (Figure 1). The initial determination of the DNA sequence (SEQ ID NO:2) of the resulting cDNA clone is shown in Figure 2. Subsequent work with this DNA fragment required confirmation of some of the features of its sequence. Approximately ten nucleotide residues were found to have been listed in error. Thus a corrected sequence for this DNA fragment is listed in SEQ ID NO:14 and the deduced amino acid sequence is set forth in SEQ ID NO:15. The revised sequences form the bases for analyses and comparisons reported herein.

EXAMPLE 2

Overexpression of the *Arabidopsis* cDNA in *E. coli*

The deduced amino acid sequence for *Arabidopsis* *p*-hydroxyphenylpyruvate dioxygenase was aligned with the amino acid sequences of *p*-hydroxyphenylpyruvate dioxygenase from mouse, pig, and *Streptomyces avermitilis* using the Pileup program of GCG (Program Manual for the Wisconsin Package, Version 8, September 1994, Genetics Computer Group, 575 Science Drive, Madison, WI, USA 53711). This analysis suggested an additional

29 amino acid-extension at the amino terminus of the *Arabidopsis* sequence (positions 1-29, Figure 3 and SEQ ID NO:3). This amino-terminal extension was assumed to be a chloroplast transit peptide which would be absent from the mature enzyme. Therefore, removal of the chloroplast transit peptide coding sequence coincided with transfer of the *p*-hydroxyphenylpyruvate dioxygenase coding sequence from the cloning vector into the expression vector.

The *Arabidopsis p*-hydroxyphenylpyruvate dioxygenase cDNA was moved from the pBluescript SK- cloning vector (Stratagene, La Jolla, CA) to the pET24c(+) expression vector (Novagen, Madison, WI) through the intermediate cloning vector pT7BlueR (Novagen). The plasmid pGBPPD2 consists of the *Arabidopsis p*-hydroxyphenylpyruvate dioxygenase cDNA and the pBluescript SK- cloning vector (Stratagene). The plasmid pE24CP1 consists of the *Arabidopsis p*-hydroxyphenylpyruvate dioxygenase cDNA, without the putative chloroplast transit peptide DNA sequence, and the pET24c(+) expression vector (Novagen).

The plasmids pGBPPD2 and pT7BlueR (5 µg each) were individually digested with 20 units of Xba I (New England Biolabs, NEB, Beverly, MA) and 20 units of Hind III (Gibco BRL, Gaithersburg, MD) in NEB restriction enzyme buffer 2 supplemented with 100 µg/mL bovine serum albumin at 37 °C for 1.75 h. Digesting pGBPPD2 with the restriction enzymes Xba I and Hind III releases the 5' and 3' ends, respectively, of the *p*-hydroxyphenylpyruvate dioxygenase cDNA from the pBluescript SK- polylinker. Products of the digestion were electrophoretically separated in a 1 percent agarose gel using TRIS/acetate/EDTA (TAE) buffer and visualized with ethidium bromide staining (Maniatis). Digestion of pGBPPD2 with the two restriction endonucleases resulted in a 2922 bp vector band and 1499 bp *p*-hydroxyphenylpyruvate dioxygenase cDNA band. Only a 2863 bp band was apparent after digesting pT7BlueR with the two enzymes, although a 24 bp fragment would also result. The 1499 bp *p*-hydroxyphenylpyruvate dioxygenase band and the 2863 bp T7BlueR band were cut out of the gel and the associated DNA purified from the agarose using a QIAquick Gel Extraction Kit (Qiagen, Chatsworth, CA) according to the manufacturer's instructions. The purified DNA samples were precipitated by the addition of sodium acetate (pH 5.2) to 0.3 M, 10 µg tRNA (added as carrier), two volumes of -20 °C ethanol and incubation at -20 °C overnight. Nucleic acid pellets were collected by centrifugation, washed with 70% ethanol and air dried. Both pellets were solubilized in 10 µL of TRIS/EDTA (TE) buffer, pH 8 (Maniatis), and then 1 µL of each sample loaded onto a 1% agarose, TAE gel in separate wells next to a well containing 4 µL of Mass Ladder (Gibco BRL). All samples were adjusted

to 10 μ L with water before loading. DNA was quantified by comparing band intensities of each sample with Mass Ladder band intensities following ethidium bromide staining and UV illumination.

Approximately 300 ng of *p*-hydroxyphenylpyruvate dioxygenase insert was
5 mixed with 300 ng of double digested pT7BlueR vector in a total volume of 7 μ L and then heated to 45 °C for 5 min followed by cooling on ice. T4 DNA ligase buffer (Gibco BRL) and 1 unit of T4 DNA ligase (Gibco BRL) were added to the cooled DNA for a total volume of 10 μ L. The ligation mix was incubated at room temperature for 4 h and then transformed into MAX Efficiency DH5 α Competent
10 Cells (Gibco BRL) of *E. coli* according to standard procedures (Maniatis). Transformed bacteria were spread onto LB agar plates supplemented with 100 μ g/mL carbenicillin and incubated overnight at 37 °C. Seventeen bacterial colonies were selected for subsequent analysis. A portion of each colony was inoculated into a separate 17x100 mm polypropylene culture tube (Falcon,
15 Lincoln Park, NJ) containing 2 mL of liquid LB media and 200 μ g/mL carbenicillin. Liquid bacteria cultures were incubated overnight at 37 °C with shaking (250 rpm). Plasmid DNA was then isolated using a QIAprep Spin Plasmid Miniprep Kit (Qiagen) according to the manufacturer's instructions. A portion (5 μ L out of 50 μ L total) of each plasmid preparation was digested with
20 10 units each of Hind III and EcoR V (Gibco BRL) in a total volume of 15 μ L with React 2 buffer (Gibco BRL) for one h. (Note: The EcoRV site in the pBluescript polylinker was destroyed during the preparation of pGBPPD2 so only the EcoRV site in the pT7BlueR polylinker would be accessible to the restriction
25 tris/borate/EDTA (TBE) buffer (Maniatis). Bands were visualized with ethidium bromide staining; 7 out of 17 samples which contained 2 bands (2837 and 1525 bp) contained the *p*-hydroxyphenylpyruvate dioxygenase insert and were designated pT7BlueR+PDO1 (see Figure 4).

In order to remove the putative chloroplast transit sequence, the remaining
30 45 μ L of each prep of pT7BlueR+PDO1 were combined into a single sample and the DNA content determined spectrophotometrically at A₂₆₀ (Maniatis). A portion (5 μ g) of pT7BlueR+PDO1 was digested with 16 units of Eco47 III (MBI Fermentas) in a total volume of 100 μ L containing buffer 0 (MBI Fermentas) at 37 °C for 2 h. The digested plasmid DNA was then precipitated with sodium
35 acetate and ethanol as above and the resulting dried nucleic acid pellet was dissolved in 60 μ L of React 2 (Gibco BRL) containing 20 units of Nde I (Gibco BRL) and incubated 2 h at 37 °C. The double digested sample was then loaded onto a 1% agarose gel in TAE and the large 4166 bp Nde I-Eco47III fragment

separated from the 196 bp fragment electrophoretically. The large fragment was cut out of the gel, purified from agarose and precipitated as above.

5 An oligonucleotide mix was prepared consisting of 100 pmoles each of oligos CAM32 and CAM33 (SEQ ID NOS:4 and 5, respectively) in a combined volume of 9.9 μ L. The two oligos complement each other to form a 3' blunt end corresponding to the 5' half of an Eco47 III restriction site and also form a 5' staggered end which corresponds to the 3' half of an Nde I restriction site.

CAM 32: (SEQ ID NO:4)

10 5'-TATGTCCAAGTTCGTAAGAAAGAATCCAAAGTCTGATAAATTCAAGGTTAAGC-3'

CAM 33: (SEQ ID NO:5)

5'-GCTTAACCTTGAATTTATCAGACTTTGGATTCTTCTTACGAACTTGGACA-3'

15 The oligo mix was heated to 90 °C for 1.5 min and then allowed to cool to room temperature over 20 min. The dried nucleic acid pellet resulting from purification of the 4166 bp Nde I-Eco47 III fragment was solubilized in 7 μ L of the cooled oligo mix and subsequently heated to 45 °C for 5 min followed by cooling on ice. Ligation of the oligos with the Nde I-Eco47 III fragment followed
20 by transformation into DH5 α was performed as above. Transformed bacterial cells were spread onto LB/carbenicillin plates and incubated at 37 °C overnight. Seventeen colonies were selected and processed to isolate plasmid DNA as above. A portion (5 out of 50 μ L) of each plasmid was double digested with 10 units each of Nde I and Hind III and the fragments separated electrophoretically on a 1%
25 agarose gel in TBE. A two band pattern corresponding to insert (1373 or 1518 bp) and vector (2844 bp) was detected. An additional double digest with 10 units each of Xba I and Hind III was performed on another 5 μ L aliquot of plasmids. When digested with Nde I and Hind III, none of the plasmids which contained the smaller insert size contained a Xba I site. The Xba I site would be eliminated if
30 the two oligos replaced the 196 bp fragment originally present in pT7Blue+PDO1. The 7 plasmid samples with the modified *p*-hydroxyphenylpyruvate dioxygenase insert were combined and designated pT7BlueR+PDO2.

The pT7BlueR+PDO2 plasmid DNA was quantified spectrophotometrically (above) and then 5 μ g was digested with 20 units each of Hind III and Nde I in
35 62 μ L of React 2 for 2 h at 37 °C. The digested sample was subsequently loaded onto a 1% agarose gel in TAE and separated electrophoretically. The 1373 bp fragment was isolated and precipitated as above. The plasmid pET24c(+) (5 μ g) was double digested with 20 units each of both Nde I and Hind III in React 2 at 37 °C for 2 h and the 5245 bp fragment then gel purified on a 1% agarose gel in

TAE and subsequently separated from agarose and precipitated as above. The dried pET24c(+) pellet was solubilized in 10 μ L TE and then 8 μ L was adjusted to a 20 μ L total volume with water, dephosphorylation buffer (Gibco BRL) and 1 unit of calf intestinal alkaline phosphatase (Gibco BRL). The sample was

5 incubated at 37 °C for 30 min and then gel purified, separated from agarose, and precipitated as above. The dried, dephosphorylated, pET24c(+) vector pellet and modified *p*-hydroxyphenylpyruvate dioxygenase insert pellet were each solubilized in 10 μ L TE and then 1 μ L of each was run on a 1% agarose TBE gel with 4 μ L of mass ladder to quantify DNA as above. One hundred nanograms of modified

10 *p*-hydroxyphenylpyruvate dioxygenase insert was mixed with 120 ng of dephosphorylated pET24c(+) vector in a total of 7 μ L volume. The mix was heated to 45 °C for 5 min and then cooled on ice. The mix was then supplemented with T4 DNA ligase buffer and 1 unit of T4 DNA ligase in a total volume of 10 μ L and the mix allowed to incubate at room temperature for 4 h. The ligation

15 mix was subsequently transformed into DH5 α , spread on LB agar supplemented with 30 μ g/mL kanamycin, and incubated overnight at 37 °C. Plasmid preparations were performed on 11 colonies as above. Plasmids were double digested with Nde I and Hind III and fragments separated electrophoretically. All plasmids had the expected 1373 bp and 5245 bp fragments. One bacteria colony

20 was selected and used to inoculate 100 mL of liquid LB supplemented with 30 μ g/mL kanamycin which was subsequently incubated at 37 °C overnight with shaking. Plasmid DNA was isolated from the resulting bacteria culture using a Qiagen Plasmid Midi Kit according to the manufacturer's instructions. A portion of the plasmid DNA (pE24CP1) was sequenced with the Sequenase Version 2.0

25 DNA Sequencing Kit (United States Biochemical, Cleveland, OH) using a biotinylated sequencing primer to the T7 promoter (United State Biochemical) according to the manufacturer's instructions for non-radioactive manual sequencing. DNA was transferred from the sequencing gel to Hybond-N+ nylon transfer membrane (Amersham, Arlington Heights, IL) by capillary action.

30 Transfer and all subsequent steps in chemiluminescent detection of DNA fragments were performed with a SEQ-Light Chemiluminescent Sequencing System kit (Tropix, Bedford, MA) according to the manufacturer's instructions. DNA sequencing verified that the plasmid contained the expected 5' sequence for the modified *p*-hydroxyphenylpyruvate dioxygenase insert where nucleotides 1-95

35 (Figure 2) were replaced with an ATG transcriptional start site. This is equivalent to amino acids 2-29 (Figure 3) being eliminated from the N-terminus of the *Arabidopsis p*-hydroxyphenylpyruvate dioxygenase amino acid sequence.

The plasmid pE24CP1 was transformed into competent cells of BL21(DE3) *E. coli* (Novagen), as above. Transformed cells were spread on LB/kanamycin plates and incubated overnight at 37 °C. Seven colonies were selected for plasmid preparations as above and plasmid DNA was double digested with Nde I and Hind III to verify that all plasmids had the expected electrophoretic banding pattern. One colony was selected and streaked for isolation on LB/kanamycin plates. A well isolated colony was used to inoculate liquid LB supplemented with 30 µg/mL kanamycin and the culture was incubated at 37 °C with shaking (250 rpm) until it reached an A₆₀₀ of 0.6 absorbance units. An 8% glycerol freezer stock was prepared according to the Novagen protocol and stored at -80 °C. All subsequent expression studies were done with freshly grown bacterial cells that were isolated from LB/kanamycin plates streaked from the glycerol freezer stock.

BL21(DE3) *E. coli* cells containing either pE24CP1 or pET24c(+) (negative control) were streaked out onto LB/kanamycin plates from a glycerol freezer stock (above) and incubated overnight at 37 °C. One isolated colony was selected for inoculation of 2 mL of LB containing 30 µg/mL kanamycin in a 17 x 100 mm Falcon tube, and the culture was incubated at 37 °C with shaking (250 rpm) overnight. The overnight cultures were then used to inoculate 100 mL of fresh LB containing 30 µg/mL kanamycin. The new cultures were incubated at 37 °C with shaking until the A₆₀₀ reached between 0.4 and 0.6 absorbance units. One half of the pE24CP1 and pET24c(+) cultures were placed in new culture flasks and IPTG (isopropylthio-β-D-galactoside; Gibco BRL) was added to the new flasks to give a final concentration of 1 mM. The flasks were incubated an additional 3 h at 37 °C with shaking, and then the cells were harvested.

The harvested cells were centrifuged and the resulting cell pellet extracted by sonication (3 x 10 sec bursts) in 2 mL extraction buffer (50 mM (20 mM in the first experiment; Table 2) potassium phosphate buffer, pH 7.2, containing 0.14 M KCl, 0.32 mM reduced glutathione, 1% polyvinylpolypyrrolidone, and 0.1% Triton X 100 (0.01% lysozyme was included in the first experiment only)). The lysate represents the crude extracted enzyme after centrifugation at 17000 g for 10 min. In the first experiment (Table 2) a 20 to 60% ammonium sulfate precipitated enzyme fraction was also assayed. Solid ammonium sulfate was slowly added with stirring to 2 mL of the lysate to bring the concentration to 20% (w/v). After incubation on ice for approximately 15 min, the solution was centrifuged at 17000 g for 10 min. The supernatant liquid was harvested and solid ammonium sulfate was added to increase the concentration to 60% (w/v). After

centrifugation, the resulting pellet was resuspended in 1 mL of the extraction buffer.

A portion of the insoluble protein resulting from expression of *Arabidopsis* *p*-hydroxyphenylpyruvate dioxygenase in bacteria was utilized for N-terminal sequence analysis. The protein (approximately 180 µg) was suspended in 60 µL of extraction buffer and then diluted with 5 volumes of sample buffer (62.5 mM Tris, pH 6.8, 6 M urea, 160 mM dithiothreitol, 0.01% bromophenol blue) followed by intermittent vortexing for one hour at room temperature. A 1.5 mm thick, 12% polyacrylamide resolving gel was prepared for a Mini-Protein II dual slab cell (Bio-Rad, Hercules, CA) using the manufacturer's instructions. The polyacrylamide was allowed to polymerize for 3 h and then a stacking gel was prepared using a preparative comb. The running buffer was prepared according to the manufacturer's instructions with the addition of 0.1 mM sodium thioglycolate. The solubilized protein sample was electrophoretically separated using the manufacturer's instructions. When the bromophenol blue dye front reached the bottom of the gel, the gel was removed and equilibrated for 5 min in blotting buffer (10 mM CAPS, pH 11, 10% methanol, balance water). The gel was then placed in a Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad), according to the manufacturer's instructions, with a ProBlott PVDF membrane (Applied Biosystems, Foster City, CA) treated according to the manufacturer's instruction. Electroblothing was done in the presence of blotting buffer at 50 volts for 45 min in an ice bath. The membrane was then rinsed in water and stained with Coomassie Blue as described in the ProBlott protocol. The major protein band was excised from the membrane and subjected to N-terminal amino acid sequencing on a Beckman (Fullerton, CA) LF3000 protein sequencer. The first 11 cycles identified S-K-F-V-R-K-N-P-K-S-D (see SEQ ID NO:3, amino acids 30-40), respectively. This is the expected N-terminus of the modified *Arabidopsis* *p*-hydroxyphenylpyruvate dioxygenase minus the initial methionine (amino acids 30-40, Figure 3).

EXAMPLE 3

p-Hydroxyphenylpyruvate Dioxygenase Enzymatic Activity of the Plant Protein Expressed in *E. Coli*

Cell cultures with different plasmid constructs were extracted as described above and assayed by measuring the formation of $^{14}\text{CO}_2$ from [1- ^{14}C]-*p*-hydroxyphenylpyruvate or $^{14}\text{CO}_2$ and ^{14}C -homogenisate from [U- ^{14}C]-*p*-hydroxyphenylpyruvate (Lindblad, B., (1971) *Clin. Chim. Acta* 34:113-121; and Lindstedt, S. and Odelhog, B., (1987) *Methods in Enzymology* 142:143-148). The labeled substrate was prepared from [1- ^{14}C]-L-tyrosine

(55 mCi/mmol; American Radiolabeled Chemicals, Inc., St. Louis, MO) or [U-¹⁴C]-L-tyrosine (498 mCi/mmol; DuPont NEN, Boston, MA). A 50-100 μ L aliquot (5-10 μ Ci) of the of the labeled tyrosine stock solution was transferred to a 4 mL glass vial and blown to dryness in a stream of nitrogen at 45°C. To the vial
5 was added 175 μ L of 0.1 M phosphate buffer, pH 6.5, 5 μ L catalase (28,700 units of C-100, Sigma Chemical Co., St. Louis, MO), and 20 μ L L-amino acid oxidase (Sigma A-9253, 6.5 units/mL). The vial was then placed on a shaker water bath set at 30°C, 60 cycles/min, for 0.5 to 1 h. The reaction mix was then passed
10 through a small column containing 400 μ L Dowex AG 50W X8 cation exchange resin. The column was then washed with 1.5 mL of water and the eluant containing the labeled *p*-hydroxyphenylpyruvate was collected. The labeled substrate was either used immediately or stored at -80°C and used within a week after preparation.

The assay was performed in 14 mL culture tubes capped with serum
15 stoppers through which a polypropylene well containing 200 μ L of 1 N KOH was suspended. The reaction mixture contained 5.740 units of catalase, 100 μ L of a freshly prepared 1:1 (v:v) mixture of 150 mM reduced glutathione and 3 mM dichlorophenolindophenol, 5 mM ascorbate, 0.1 mM ferrous sulfate (the ascorbate and ferrous sulfate were not present in the buffer used in the first experiment;
20 Table 2), 50 μ M unlabeled *p*-hydroxyphenylpyruvate, 1-25 μ L of the enzyme extract, and 50 mM potassium phosphate buffer in a final volume of 980 μ L. Unlabeled substrate was made fresh daily in 50 mM potassium phosphate buffer and allowed to equilibrate for at least 2 h at room temperature to insure that
25 greater than 95% was in the keto form. The tubes were incubated for 10 min at 30°C in a shaking water bath prior to adding 20 μ L (0.04 μ Ci) of ¹⁴C-*p*-hydroxyphenylpyruvate. The reaction was terminated after 60 min by injecting 500 μ L of 1 N sulfuric acid through the serum stopper. The vials were left on the shaker for another 30 min to insure complete capture of the released ¹⁴CO₂. The serum caps were then removed and the wells cut and dropped into
30 8 mL scintillation vials. Six mL of Formula-989 scintillation fluid (Packard Instruments, Meriden, CT) was added to the vials and the ¹⁴C radioactivity was determined by scintillation counting. Table 2 summarizes the results of this experiment.

Table 2

p-Hydroxyphenylpyruvate Dioxygenase Activity of Extracts from
E. coli Containing Different Plasmid Constructs

| Plasmid | Inducer (1 mM IPTG) | Lysate | | Ammonium Sulfate Precipitate | |
|-----------|------------------------|-----------|---------------|------------------------------|---------------|
| | | dpm * /mg | nmol/min x mg | dpm * /mg | nmol/min x mg |
| pET24c(+) | - | 12,318 | 0.09 | 0 | 0.00 |
| pET24c(+) | + | 35,115 | 0.25 | 3,393 | 0.03 |
| pE24CPI | - | 24,607 | 0.17 | 126,761 | 0.89 |
| pE24CPI | + | 243,801 | 1.71 | 1,371,823 | 9.64 |

* $^{14}\text{C} : ^{12}\text{C} = 1 : 50$; sp. act. of ^{14}C -*p*-hydroxyphenylpyruvate = 55 mCi/mmol

5

The results show there was little or no *p*-hydroxyphenylpyruvate dioxygenase activity in any of the cell cultures that did not have the plasmid containing the nucleic acid fragment encoding *p*-hydroxyphenylpyruvate dioxygenase (pET24c(+)) and the inducer of gene expression (IPTG). The gene and inducer together resulted in a marked increase in activity.

10

In the experiment with $[\text{U-}^{14}\text{C}]$ *p*-hydroxyphenylpyruvate ("HPPA"), where both $^{14}\text{CO}_2$ and ^{14}C -homogentisic acid were measured, the reaction was initiated by adding 50 μL of labeled substrate (0.3 μCi) and was terminated with 100 μL of 10% phosphoric acid. The $^{14}\text{CO}_2$ released was determined by scintillation counting, while the level of homogentisic acid was determined by HPLC on a Zorbax RX-C8 column (4.6 x 250 mm) with an in-line radioactivity detector. Aliquots of 1.7 to 15 μL were taken from the reaction mix after centrifugation and diluted into the column equilibration buffer prior to injection. Separation was performed at ambient temperature with a flow rate of 1.0 mL/min and the following gradient with solvent A and B being water and methanol, each with 1% phosphoric acid: 0-2 min, isocratic at 95% A and 5% B; 2-17 min. linear gradient from 95 to 75% A and 5 to 25% B; 17-19 min linear gradient from 75 to 5% A and 25 to 95% B; 19-22 min, isocratic at 5% A and 95% B; 22-24 min, linear gradient from 5% to 95% A and 95 to 5% B. In this system homogentisate eluted at 10.8 min. The results from this experiment are shown in Table 3.

15

20

25

Table 3

p-Hydroxyphenylpyruvate Dioxygenase Activity of Cell Extracts
Determined by CO₂ Release and Homogentisic Acid Synthesis
from [U-¹⁴C] *p*-Hydroxyphenylpyruvate

| Plasmid | Inducer (1 mM IPTG) | nmol/min x mg* | |
|-----------|------------------------|-------------------------------|-------------------|
| | | ¹⁴ CO ₂ | Homogentisic acid |
| pET24c(+) | - | 0.00 | 0.00 |
| pET24c(+) | + | 0.19 | 0.00 |
| pE24CP1 | - | 4.68 | 4.76 |
| pE24CP1 | + | 29.12 | 29.82 |

* ¹⁴C : ¹²C = 1 : 87.7; sp. act. of ¹⁴C[U]-*p*-HPPA = 498 mCi /mmol

There was a tight correlation between the results from the assays of the two products of the reaction. The results confirmed there was no significant *p*-hydroxyphenylpyruvate dioxygenase activity in either cell culture that did not contain the nucleic acid fragment encoding *p*-hydroxyphenylpyruvate dioxygenase. There was measurable enzyme activity in the absence of the inducer, but when the inducer was added the activity increased greater than six-fold over uninduced cultures. These results and those of Table 2 clearly show that the nucleic acid fragment isolated and overexpressed in *E. coli* cells encodes a protein that catalyzes the conversion of *p*-hydroxyphenylpyruvate to homogentisate with the release of CO₂.

The overexpressed protein was also assayed spectrophotometrically at ambient temperature using the enol borate-tautomerase assay (Lin, E. C. C. et al., (1958) *J. Biol. Chem.* 233:668-673). The assay buffer contained 0.4 M borate (adjusted to pH 7.2 with 0.2 M sodium borate), 4 mM ascorbate, 2.5 mM EDTA, 40 μM *p*-hydroxyphenylpyruvate, and 0.5 units of tautomerase (Sigma T-6004) per 10 mL buffer. The reaction mix was used when the tautomerization of the substrate was complete (when absorbance at 308 nm had stabilized). The assay was initiated by adding 40 μL of the cell extracts to 960 μL of the assay buffer, and the reaction was followed by measuring the decrease in absorbance at 308 nm. Table 4 summarizes the results with extracts of the same four cell cultures described in Table 3.

Table 4
Spectrophotometric Assay of *p*-Hydroxyphenylpyruvate
Dioxygenase Activity of Cell Extracts

| Plasmid | Inducer (1 mM IPTG) | nmol <i>p</i> -HP lost/min x mg* |
|-----------|------------------------|----------------------------------|
| pET24c(+) | - | 1.58 |
| pET24c(+) | + | 2.73 |
| pE24CPI | - | 4.91 |
| pE24CPI | + | 22.32 |

* Loss of *p*-hydroxyphenylpyruvate based on a molar extinction coefficient for the equilibrium mixture of 9850 as reported by Lin et al. ((1958) *J. Biol. Chem.* 233: 668-673).

EXAMPLE 4

Inhibition of *p*-Hydroxyphenylpyruvate Dioxygenase by Commercial Herbicides

The enzymatic activity of the overexpressed protein is inhibited by two herbicides known to inhibit plant *p*-hydroxyphenylpyruvate dioxygenase: Sulcotrione (2-(2-chloro-4-methanesulfonylbenzoyl)-1,3-cyclohexanedione); and Isoxaflutole (5-cyclopropylisoxazol-4-yl 2-mesyl-4-trifluoromethylphenyl ketone). These two compounds were tested against the overexpressed protein using both the $^{14}\text{CO}_2$ and the continuous spectrophotometric enol borate-tautomerase assays. Both compounds were added to the assay buffers in 10 μL of acetone or dimethyl sulfoxide. The I_{50} values (concentration inhibiting the enzyme 50%) were calculated based on the percent inhibition observed over several concentrations of the inhibitor. The results of the assays are shown in Table 5.

Table 5
 I_{50} Values of Inhibitors of Plant *p*-Hydroxyphenylpyruvate Dioxygenase

| Compound | I_{50} value (nM) derived from | |
|--------------|----------------------------------|--------------------------|
| | $^{14}\text{CO}_2$ assay | spectrophotometric assay |
| sulcotrione | 43 | 44 |
| isoxaflutole | 409 | 1042 |

These results clearly show that the *p*-hydroxyphenylpyruvate dioxygenase activity of the overexpressed protein is inhibited by commercial herbicides that have inhibition of this enzyme as their mode of action. Moreover, the continuous spectrophotometric assay gave similar I_{50} values to those obtained with the $^{14}\text{CO}_2$ assay. The spectrophotometric assay can be adapted to a high capacity screen for

inhibitors of *p*-hydroxyphenylpyruvate dioxygenase by adapting it to a microtiter plate assay combined with a plate reader that would read at or near 308 nm. Furthermore, any colorimetric or fluorescent assay for homogentisate or *p*-hydroxyphenylpyruvate would also be able to be readily adapted into a high capacity screen for inhibitors of this enzyme. The isolated overexpressed enzyme has sufficient activity to be used directly in a spectrophotometric assay or it can be further purified for enhanced assay sensitivity.

EXAMPLE 5

Re-construction of the Full-length *p*-Hydroxyphenylpyruvate Dioxygenase Gene for Production of Active, Stable Enzyme in Bacteria

The plasmid pT7BlueR+PDO2, described in Example 2 and containing the full-length *p*-hydroxyphenylpyruvate dioxygenase gene, proved to have incorrect sequence at the EcoRI site. This was re-sequenced so that an oligonucleotide could be designed to replace the EcoRI site with an NdeI site using conventional loop-out mutagenesis. The oligonucleotide was designed so that this procedure also introduced an ATG initiation codon at the 5'-end of the *p*-hydroxyphenylpyruvate dioxygenase gene followed by the full-length *p*-hydroxyphenylpyruvate dioxygenase sequence. After mutagenesis, the clone was amplified in *E. coli* and the plasmid was purified. The resulting full-length gene, "PDO-B", was then digested with the enzymes using NdeI and NheI, and the ~820 bp fragment used to replace the NdeI - NheI segment of the truncated *p*-hydroxyphenylpyruvate dioxygenase gene, "PDO-A," in pE24CP1 (Example 1). The resulting plasmid, pE24PDO-B can be expressed in bacteria to produce the full-length *Arabidopsis p*-hydroxyphenylpyruvate dioxygenase enzyme as determined by enzyme activity and N-terminal sequence analysis.

EXAMPLE 6

Enhanced Stability of Full Length Construct Over the Truncated Construct

Two different constructs for *Arabidopsis thaliana p*-hydroxyphenylpyruvate dioxygenase, one containing the full-length sequence, PDO-B as described in Example 5 and produced from plasmid pE24PDO-B, and one containing the truncated sequence lacking the putative chloroplast leader sequence, PDO-A as produced from plasmid pE24CP1, were both purified to the same extent using a Pharmacia phenyl Sepharose hydrophobic interaction column followed by gel filtration chromatography on Pharmacia Sephacryl 300. The two proteins were diluted to 1 mg/mL in 20 mM bis tris-propane buffer, pH 7.2 containing 5 mM ascorbate, 1 mM reduced glutathione and 0.1 mM ferrous ammonium sulfate and stored in a refrigerator at 4 °C for up to 10 days. Aliquots were removed at various times and assayed for activity using the tautomerase

coupled spectrophotometric assay. Under these conditions the half-life for the activity of the full length enzyme was 4 days, whereas the truncated enzyme preparation had a half-life of 9 to 10 hours. In addition, the activity of the full length enzyme could be restored by incubation with iron and reducing agent, reduced glutathione or ascorbate, or by dialysis against buffer containing iron and reducing agent. In contrast, the activity of the truncated enzyme could not be restored by incubation with or dialysis against buffer containing iron and reducing agent. The full-length enzyme was also more stable in the spectrophotometric assay showing a 2 to 3 times longer useful linear region than the truncated enzyme. Both enzyme preparations showed similar I_{50} values with the herbicidally active inhibitors.

These results clearly show that the full-length PDO-B construct has decided advantages over the truncated enzyme due to the enhanced stability under storage conditions, in the spectrophotometric assay and in the reversible reconstitution of activity in the presence of iron and reducing agent. While both enzyme constructs can be used for screening of inhibitors, the PDO-B enzyme is preferred for this application and is far superior for mechanistic and structural studies.

EXAMPLE 7

Cloning of the Maize *p*-Hydroxyphenylpyruvate Dioxygenase Gene

Approximately 600,000 plaques of a Stratagene maize Uni-Zap cDNA library (from young plants) were screened by filter hybridization under moderate stringency using a heterologous probe. The probe was prepared by PCR and was a 916 bp fragment of DNA having the sequence defined by the region extending from position 263 to 1178 of SEQ ID NO:14. Twenty-four positive phage clones were identified in the primary screen, and eleven phage clones were recovered from a secondary screen. Seven positive clones were submitted for sequencing, and four showed significant conservation sequence at the amino acid level when compared with the *Arabidopsis thaliana* *p*-hydroxyphenylpyruvate dioxygenase protein. The longest of the four contained an insert of 988 bp and showed 70% identity and 78% similarity with the *Arabidopsis* protein, but was lacking approximately 550 bp corresponding to the amino terminal end of the protein.

Attempts to obtain a full-length cDNA of the maize *p*-hydroxyphenylpyruvate dioxygenase gene were unsuccessful, possibly because the secondary structure of the RNA inhibited efficient reverse transcription of this transcript. Two additional cDNA libraries were screened and clones long enough to contain a full-length cDNA were sequenced. All of these clones were shown to be chimeras. Therefore a genomic library was screened to obtain the 5' one-third of

the gene. Approximately 1 million clones from a Clontech *Zea mays* (var. B73) library in the phage vector EMBL3 (whole seedlings, 2 leaf stage) were screened using a 415 bp EcoRI-BssHII fragment containing the 5' end of the truncated corn *p*-hydroxyphenylpyruvate dioxygenase cDNA (clone H1011C). Eight positive
5 primary phage clones were plated and screened, and four secondary clones were picked. DNA was prepared from each using the Qiagen Lambda midi-kit. Restriction digests with Sall or EcoRI indicated that two clones were the same. DNA samples from the remaining 3 clones (11.1.3, 13.1.1, and 21.2.1) were
10 digested with Sall, EcoRI, or Sall and EcoRI, prepared for Southern analysis, and probed with the full length *Arabidopsis p*-hydroxyphenylpyruvate dioxygenase gene. Two of the clones (11.1.3 and 13.1.1) showed sequence conservation, and these homologous fragments were subcloned and sequenced. Both clones
appeared to contain the full-length gene and each contained one intron near the 3' end of the gene. However, there were differences between the sequences of the
15 two clones indicating that they may be two different genes or one may be a pseudogene. The sequence of clone 11.1.3 matched the cDNA sequence, and this clone was used to construct a full length *p*-hydroxyphenylpyruvate dioxygenase coding region.

The gene was contained on two adjacent fragments, a 3.5 kb EcoRI - Sall
20 fragment and a 2 kb Sall fragment. Both were subcloned into pBluescript SKII+ resulting in the plasmids pES1113 and pSal1113. pES1113 was digested with SpeI to release approximately 2.7 kb of upstream sequence and then religated, resulting in a plasmid with an insert of 747 base pairs (pSPE1). pSPE1 was
digested with Sall to linearize the plasmid and ligated with the 2 kb Sall fragment
25 from pSal1113, which had been released by digestion with Sall and gel purified. Orientation was confirmed by digestion with SpeI and Bpu1102I and the correct plasmid was named p1113. In order to remove the intron contained in the 3' end of the genomic clone, the plasmid was digested with Bpu1102I and XhoI and the
3.9 kb fragment containing the vector and 5' part of the gene was gel purified.
30 The corresponding 882 bp Bpu1102I-XhoI fragment from pH1011c (cDNA) was gel purified and ligated with this 3.9 kb fragment resulting in the clone pMPDO (ATCC 209120), which contains a 1782 bp insert. There are 260 base pairs upstream of the putative ATG and 189 base pairs downstream of the stop codon. The full-length sequence was confirmed by sequencing across the insert. The
35 nucleic acid sequence and the deduced protein sequence for corn *p*-hydroxyphenylpyruvate dioxygenase are presented in SEQ ID NOS:10 and 11, respectively. The sequences for *p*-hydroxyphenylpyruvate dioxygenases obtained from corn and *Arabidopsis* were compared using the "Gap" program of GCG

(Program Manual for the Wisconsin Package, Version 9.0-OpenVMS. December 1996, Genetics Computer Group, 575 Science Drive, Madison, WI, USA 53711).

The results of these comparisons indicated that these functions are approximately 67% identical at the nucleotide level, and they possess 69% similarity and 62%

5 identity at the amino acid level. The predicted amino acid sequence of corn *p*-hydroxyphenylpyruvate dioxygenase is compared with that from *Arabidopsis* and other eukaryotes in Figure 3.

EXAMPLE 8

Composition of a cDNA Library; Isolation and Sequencing of cDNA Clones

10 A cDNA library representing mRNAs from developing seeds of *Vernonia galamenensis* that had just begun production of vernolic acid was prepared. The library was prepared in a Uni-ZAP™ XR vector according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). Conversion of the Uni-ZAP™ XR library into a plasmid library was accomplished according to the
15 protocol provided by Stratagene. Upon conversion, cDNA inserts were contained in the plasmid vector pBluescript. cDNA inserts from randomly picked bacterial colonies containing recombinant pBluescript plasmids were amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences. Amplified insert DNAs were sequenced in dye-
20 primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams, M. D. et al., (1991) *Science* 252:1651). The resulting ESTs were analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

EXAMPLE 9

Identification and Characterization of cDNA Clones

25 ESTs encoding *Vernonia galamenensis* enzymes were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F. et al., (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database
30 (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 9 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database
35 using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish, W. and States, D. J.

(1993) *Nature Genetics* 3:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

The BLASTX search using clone vs1.pk0015.b2 revealed similarity of the protein encoded by the cDNA to a number of *p*-hydroxyphenylpyruvate dioxygenases from sources other than plants. The three most similar *p*-hydroxyphenylpyruvate dioxygenase proteins were a streptomycete *p*-hydroxyphenylpyruvate dioxygenase (GenBank Accession No. U11864; pLog = 8.34), a rat *p*-hydroxyphenylpyruvate dioxygenase (GenBank Accession No. M18405; pLog = 7.66), and a human *p*-hydroxyphenylpyruvate dioxygenase (GenBank Accession No. U29895; pLog = 7.60). SEQ ID NO:16 shows the nucleotide sequence of a portion of the *Vernonia galamenensis* cDNA in clone vs1.pk0015.b2. Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragment encodes a portion of *Vernonia galamenensis* *p*-hydroxyphenylpyruvate dioxygenase.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT:
 - (A) NAME: E. I. DUPONT DE NEMOURS AND COMPANY
 - (B) STREET: 1007 MARKET STREET
 - (C) CITY: WILMINGTON
 - (D) STATE: DELAWARE
 - (E) COUNTRY: U.S.A.
 - (F) POSTAL CODE (ZIP): 19898
 - (G) TELEPHONE: 302-892-8112
 - (H) TELEFAX: 302-773-0164
 - (I) TELEX: 6717325
- (ii) TITLE OF INVENTION: PLANT GENE FOR *p*-HYDROXY-PHENYLPYRUVATE DIOXYGENASE
- (iii) NUMBER OF SEQUENCES: 16
- (iv) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: DISKETTE, 3.50 INCH
 - (B) COMPUTER: IBM PC COMPATIBLE
 - (C) OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS
 - (D) SOFTWARE: MICROSOFT WORD VERSION 7.0A
- (v) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (vi) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: 60/021,364
 - (B) FILING DATE: JUNE 27, 1996
- (vii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: FLOYD, LINDA AXAMETHY
 - (B) REGISTRATION NUMBER: 33,692
 - (C) REFERENCE/DOCKET NUMBER: BA-9120

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 233 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CAAGAAACGN GTCGNCGACG TGCTCAGCGA TGATCAGATC AAGGAGTGTG AGGAATTAGG 60
 GATTCTTNTA GACAGAGATG ATCAAGGGAC GTTNCCTCAA ATCTNCACAA AACCCTAGG 120
 TGACAGGCCG ACGNTATTTA TAGAGATAAT CCAGAGNCTA GGATGCATGA TGAAAGATGT 180
 GGAAGGGGANG GCTTACCAGA GTGGAGNATN TNGTGTTTT GGCAAAGGCA ATT 233

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1448 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 9..1343

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TGAAATCA ATG GGC CAC CAA AAC GCC GCC GTT TCA GAG AAT CAA AAC CAT 50
 Met Gly His Gln Asn Ala Ala Val Ser Glu Asn Gln Asn His
 1 5 10

GAT GAC GGC GCT GCG TCG TCG CCG GGA TTC AAG CTC GTC GGA TTT TCC 98
 Asp Asp Gly Ala Ala Ser Ser Pro Gly Phe Lys Leu Val Gly Phe Ser
 15 20 25 30

AAG TTC GTA AGA AAG AAT CCA AAG TCT GAT AAA TTC AAG GTT AAG CGC 146
 Lys Phe Val Arg Lys Asn Pro Lys Ser Asp Lys Phe Lys Val Lys Arg
 35 40 45

TTC CAT CAC ATC GAG TTC TGG TGC GGG GAC GCA ACC AAC GTC GCT CGT 194
 Phe His His Ile Glu Phe Trp Cys Gly Asp Ala Thr Asn Val Ala Arg
 50 55 60

CGC TTC TCC TGG GGT CTG GGG ATG AGA TTC TCC GCC AAA TCC GAT CTT 242
 Arg Phe Ser Trp Gly Leu Gly Met Arg Phe Ser Ala Lys Ser Asp Leu
 65 70 75

TCC ACC GGA AAC ATG GTT CAC GCC TCT TAC CTA CTC ACC TCC GGT GAA 290
 Ser Thr Gly Asn Met Val His Ala Ser Tyr Leu Leu Thr Ser Gly Glu
 80 85 90

CTC CGA TTC CTT TTC ACT GCT CCT TAC TCT CCG TCT CTC TCC GGC GGA 338
 Leu Arg Phe Leu Phe Thr Ala Pro Tyr Ser Pro Ser Leu Ser Gly Gly
 95 100 105 110

| | |
|---|------|
| GAG ATT AAA CCG ACA ACC ACA GGT TCT ATC CCA AGT TTC GAT CAC GGG Glu Ile Lys Pro Thr Thr Thr Gly Ser Ile Pro Ser Phe Asp His Gly 115 120 125 | 386 |
| TCT TGT CGG TCC TTC TTC TCT TCA CAT GGT CTC GGT GTT AGA CCC GTT Ser Cys Arg Ser Phe Phe Ser Ser His Gly Leu Gly Val Arg Pro Val 130 135 140 | 434 |
| GCG ATT GAA GTA GAA GAC GCG GAG TCA GCT TTC TCC ATC AGT GTA GCT Ala Ile Glu Val Glu Asp Ala Glu Ser Ala Phe Ser Ile Ser Val Ala 145 150 155 | 482 |
| AAT GGC GCT ATT CCT TCG TCG CCT CCT ATC GTC CTC AAT GAA GCA GTT Asn Gly Ala Ile Pro Ser Ser Pro Pro Ile Val Leu Asn Glu Ala Val 160 165 170 | 530 |
| ACG ATC GCT GAG GTT AAA CTA TAC GGC GAT GTT GTT CTC CGA TAT GTT Thr Ile Ala Glu Val Lys Leu Tyr Gly Asp Val Val Leu Arg Tyr Val 175 180 185 190 | 578 |
| AGT TAC AAA GCA GAA GAT ACC GAA AAA TCC GAA TTC TTG CCA GGG TTC Ser Tyr Lys Ala Glu Asp Thr Glu Lys Ser Glu Phe Leu Pro Gly Phe 195 200 205 | 626 |
| GAG CGT GTA GAG GAT GCG TCG TCG TTC CCA TTG GAT TAT GGT ATC CGG Glu Arg Val Glu Asp Ala Ser Ser Phe Pro Leu Asp Tyr Gly Ile Arg 210 215 220 | 674 |
| CGG CTT GAC CAC GCC GTG GGA AAC GTT CCT GAG CTT GGT CCG GCT TTA Arg Leu Asp His Ala Val Gly Asn Val Pro Glu Leu Gly Pro Ala Leu 225 230 235 | 722 |
| ACT TAT GTA GCG GGG TTC ACT GGT TTT CAC CAA TTC GCA GAG TTC ACA Thr Tyr Val Ala Gly Phe Thr Gly Phe His Gln Phe Ala Glu Phe Thr 240 245 250 | 770 |
| GCA GAC GAC GTT GGA ACC GCC GAG AGC GGT TTA AAT TCA GCG GTC CTG Ala Asp Asp Val Gly Thr Ala Glu Ser Gly Leu Asn Ser Ala Val Leu 255 260 265 270 | 818 |
| GCT AGC AAT GAT GAA ATG GTT CTT CTA CCG ATT AAC GAG CCA GTG CAC Ala Ser Asn Asp Glu Met Val Leu Leu Pro Ile Asn Glu Pro Val His 275 280 285 | 866 |
| GGA ACA AAG AAG AAG AGT CAG ATT CAG ACG TAT TTG GAA CAT AAC GAA Gly Thr Lys Arg Lys Ser Gln Ile Gln Thr Tyr Leu Glu His Asn Glu 290 295 300 | 914 |
| GGC GCA GGG CTA CAA CAT CTG GCT CTG ATG AGT GAA GAC ATA TTC AGG Gly Ala Gly Leu Gln His Leu Ala Leu Met Ser Glu Asp Ile Phe Arg 305 310 315 | 962 |
| ACC CTG AGA GAG ATG AGG AAG AGG AGC AGT ATT GGA GGA TTC GAC TTC Thr Leu Arg Glu Met Arg Lys Arg Ser Ser Ile Gly Gly Phe Asp Phe 320 325 330 | 1010 |
| ATG CCT TCT CCT CCG CCT ACT TAC TAC CAG AAT CTC AAG AAA CGG GTC Met Pro Ser Pro Pro Pro Thr Tyr Tyr Gln Asn Leu Lys Lys Arg Val 335 340 345 350 | 1058 |
| GGC GAC GTG CTC AGC GAT GAT CAG ATC AAG GAG TGT GAG GAA TTA GGG Gly Asp Val Leu Ser Asp Asp Gln Ile Lys Glu Cys Glu Glu Leu Gly 355 360 365 | 1106 |

ATT CTT GTA GAC AGA GAT GAT CAA GGG ACG TTG CTT CAA ATC TTC ACA 1154
 Ile Leu Val Asp Arg Asp Asp Gln Gly Thr Leu Leu Gln Ile Phe Thr
 370 375 380

AAA CCA CTA GGT GAC AGG CCG ACG ATA TTT ATA GAG ATA ATC CAG AGA 1202
 Lys Pro Leu Gly Asp Arg Pro Thr Ile Phe Ile Glu Ile Ile Gln Arg
 385 390 395

GTA GGA TGC ATG ATG AAA GAT GAG GAA GGG AAG GCT TAC CAG AGT GGA 1250
 Val Gly Cys Met Met Lys Asp Glu Glu Gly Lys Ala Tyr Gln Ser Gly
 400 405 410

GGA TGT GGT GGT TTT GCC AAA GGC AAT TTC TCT GAG CTC TTC AAG TCC 1298
 Gly Cys Gly Gly Phe Ala Lys Gly Asn Phe Ser Glu Leu Phe Lys Ser
 415 420 425 430

ATT GAA GAA TAC GAA AAG ACT CTT GAA GCC AAA CAG TTA GTG GGA 1343
 Ile Glu Glu Tyr Glu Lys Thr Leu Glu Ala Lys Gln Leu Val Gly
 435 440 445

TGAAACAAGAA GAAGAACCAA CTAAAGGATT GTGTAATTAA TGTAAACTG TTTTCTCTTA 1403

TCAAAACAAT GTATACAACA TCTCATTTAA AAACGAGATC AATCC 1448

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 445 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Gly His Gln Asn Ala Ala Val Ser Glu Asn Gln Asn His Asp Asp
 1 5 10 15

Gly Ala Ala Ser Ser Pro Gly Phe Lys Leu Val Gly Phe Ser Lys Phe
 20 25 30

Val Arg Lys Asn Pro Lys Ser Asp Lys Phe Lys Val Lys Arg Phe His
 35 40 45

His Ile Glu Phe Trp Cys Gly Asp Ala Thr Asn Val Ala Arg Arg Phe
 50 55 60

Ser Trp Gly Leu Gly Met Arg Phe Ser Ala Lys Ser Asp Leu Ser Thr
 65 70 75 80

Gly Asn Met Val His Ala Ser Tyr Leu Leu Thr Ser Gly Glu Leu Arg
 85 90 95

Phe Leu Phe Thr Ala Pro Tyr Ser Pro Ser Leu Ser Gly Gly Glu Ile
 100 105 110

Lys Pro Thr Thr Thr Gly Ser Ile Pro Ser Phe Asp His Gly Ser Cys
 115 120 125

Arg Ser Phe Phe Ser Ser His Gly Leu Gly Val Arg Pro Val Ala Ile
 130 135 140

Glu Val Glu Asp Ala Glu Ser Ala Phe Ser Ile Ser Val Ala Asn Gly
 145 150 155 160

Ala Ile Pro Ser Ser Pro Pro Ile Val Leu Asn Glu Ala Val Thr Ile
 165 170 175
 Ala Glu Val Lys Leu Tyr Gly Asp Val Val Leu Arg Tyr Val Ser Tyr
 180 185 190
 Lys Ala Glu Asp Thr Glu Lys Ser Glu Phe Leu Pro Gly Phe Glu Arg
 195 200 205
 Val Glu Asp Ala Ser Ser Phe Pro Leu Asp Tyr Gly Ile Arg Arg Leu
 210 215 220
 Asp His Ala Val Gly Asn Val Pro Glu Leu Gly Pro Ala Leu Thr Tyr
 225 230 235 240
 Val Ala Gly Phe Thr Gly Phe His Gln Phe Ala Glu Phe Thr Ala Asp
 245 250 255
 Asp Val Gly Thr Ala Glu Ser Gly Leu Asn Ser Ala Val Leu Ala Ser
 260 265 270
 Asn Asp Glu Met Val Leu Leu Pro Ile Asn Glu Pro Val His Gly Thr
 275 280 285
 Lys Arg Lys Ser Gln Ile Gln Thr Tyr Leu Glu His Asn Glu Gly Ala
 290 295 300
 Gly Leu Gln His Leu Ala Leu Met Met Glu Asp Ile Phe Arg Thr Leu
 305 310 315 320
 Arg Glu Met Arg Lys Arg Ser Ser Ile Gly Gly Phe Asp Phe Met Pro
 325 330 335
 Ser Pro Pro Pro Thr Tyr Tyr Gln Asn Leu Lys Lys Arg Val Gly Asp
 340 345 350
 Val Leu Ser Asp Asp Gln Ile Lys Glu Cys Glu Glu Leu Gly Ile Leu
 355 360 365
 Val Asp Arg Asp Asp Gln Gly Thr Leu Leu Gln Ile Phe Thr Lys Pro
 370 375 380
 Leu Gly Asp Arg Pro Thr Ile Phe Ile Glu Ile Ile Gln Arg Val Gly
 385 390 395 400
 Cys Met Met Lys Asp Glu Glu Gly Lys Ala Tyr Gln Ser Gly Gly Cys
 405 410 415
 Gly Gly Phe Ala Lys Gly Asn Phe Ser Glu Leu Phe Lys Ser Ile Glu
 420 425 430
 Glu Tyr Glu Lys Thr Leu Glu Ala Lys Gln Leu Val Gly
 435 440 445

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 53 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TATGTCCAAG TTCGTAAGAA AGAATCCAAA GTCTGATAAA TTCAAGGTTA AGC 53

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 51 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GCTTAACCTT GAATTTATCA GACTTTGGAT TCTTTCTTAC GAACTTGGAC A 51

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 392 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Thr Ser Tyr Ser Asp Lys Gly Glu Lys Pro Glu Arg Gly Arg Phe Leu
 1 5 10 15
 His Phe His Ser Val Thr Phe Trp Val Gly Asn Ala Lys Gln Ala Ala
 20 25 30
 Ser Tyr Tyr Cys Ser Lys Ile Gly Phe Glu Pro Leu Ala Tyr Lys Gly
 35 40 45
 Leu Glu Thr Gly Ser Arg Glu Val Val Ser His Val Val Lys Gln Asp
 50 55 60
 Lys Ile Val Phe Val Phe Ser Ser Ala Leu Asn Pro Trp Asn Lys Glu
 65 70 75 80
 Met Gly Asp His Leu Val Lys His Gly Asp Gly Val Lys Asp Ile Ala
 85 90 95
 Phe Glu Val Glu Asp Cys Asp Tyr Ile Val Gln Lys Ala Arg Glu Arg
 100 105 110
 Gly Ala Ile Ile Val Arg Glu Glu Val Cys Cys Ala Ala Asp Val Arg
 115 120 125
 Gly His His Thr Pro Leu Asp Arg Ala Arg Gln Val Trp Glu Gly Thr
 130 135 140
 Leu Val Glu Lys Met Thr Phe Cys Leu Asp Ser Arg Pro Gln Pro Ser
 145 150 155 160
 Gln Thr Leu Leu His Arg Leu Leu Leu Ser Lys Leu Pro Lys Cys Gly
 165 170 175
 Leu Glu Ile Ile Asp His Ile Val Gly Asn Gln Pro Asp Gln Glu Met
 180 185 190

Glu Ser Ala Ser Gln Trp Tyr Met Arg Asn Leu Gln Phe His Arg Phe
 195 200 205
 Trp Ser Val Asp Asp Thr Gln Ile His Thr Glu Tyr Ser Ala Leu Arg
 210 215 220
 Ser Val Val Met Ala Asn Tyr Glu Glu Ser Ile Lys Met Pro Ile Asn
 225 230 235 240
 Glu Pro Ala Pro Gly Lys Lys Lys Ser Gln Ile Gln Glu Tyr Val Asp
 245 250 255
 Tyr Asn Gly Gly Ala Gly Val Gln His Ile Ala Leu Lys Thr Glu Asp
 260 265 270
 Ile Ile Thr Ala Ile Arg Ser Leu Arg Glu Arg Gly Val Glu Phe Leu
 275 280 285
 Ala Val Pro Phe Thr Tyr Tyr Lys Gln Leu Gln Glu Lys Leu Lys Ser
 290 295 300
 Ala Lys Ile Arg Val Lys Glu Ser Ile Asp Val Leu Glu Glu Leu Lys
 305 310 315 320
 Ile Leu Val Asp Tyr Asp Glu Lys Gly Tyr Leu Leu Gln Ile Phe Thr
 325 330 335
 Lys Pro Met Gln Asp Arg Pro Thr Val Phe Leu Glu Val Ile Gln Arg
 340 345 350
 Asn Asn His Gln Gly Phe Gly Ala Gly Asn Phe Asn Ser Leu Phe Lys
 355 360 365
 Ala Phe Glu Glu Glu Gln Glu Leu Arg Gly Asn Leu Thr Asp Thr Asp
 370 375 380
 Pro Asn Gly Val Pro Phe Arg Leu
 385 390

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 392 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Thr Ser Tyr Ser Asp Lys Gly Glu Lys Pro Glu Arg Gly Arg Phe Leu
 1 5 10 15
 His Phe His Ser Val Thr Phe Trp Val Gly Asn Ala Lys Gln Ala Ala
 20 25 30
 Ser Tyr Tyr Cys Ser Lys Ile Gly Phe Glu Pro Leu Ala Tyr Lys Gly
 35 40 45
 Leu Glu Thr Gly Ser Arg Glu Val Val Ser His Val Val Lys Gln Asp
 50 55 60
 Lys Ile Val Phe Val Phe Ser Ser Ala Leu Asn Pro Trp Asn Lys Glu
 65 70 75 80

Met Gly Asp His Leu Val Lys His Gly Asp Gly Val Lys Asp Ile Ala
 85 90 95
 Phe Glu Val Glu Asp Cys Asp Tyr Ile Val Gln Lys Ala Arg Glu Arg
 100 105 110
 Gly Ala Ile Ile Val Arg Glu Glu Val Cys Cys Ala Ala Asp Val Arg
 115 120 125
 Gly His His Thr Pro Leu Asp Arg Ala Arg Gln Val Trp Glu Gly Thr
 130 135 140
 Leu Val Glu Lys Met Thr Phe Cys Leu Asp Ser Arg Pro Gln Pro Ser
 145 150 155 160
 Gln Thr Leu Leu His Arg Leu Leu Leu Ser Lys Leu Pro Lys Cys Gly
 165 170 175
 Leu Glu Ile Ile Asp His Ile Val Gly Asn Gln Pro Asp Gln Glu Met
 180 185 190
 Glu Ser Ala Ser Gln Trp Tyr Met Arg Asn Leu Gln Phe His Arg Phe
 195 200 205
 Trp Ser Val Asp Asp Thr Gln Ile His Thr Glu Tyr Ser Ala Leu Arg
 210 215 220
 Ser Val Val Met Ala Asn Tyr Glu Glu Ser Ile Lys Met Pro Ile Asn
 225 230 235 240
 Glu Pro Ala Pro Gly Lys Lys Lys Ser Gln Ile Gln Glu Tyr Val Asp
 245 250 255
 Tyr Asn Gly Gly Ala Gly Val Gln His Ile Ala Leu Lys Thr Glu Asp
 260 265 270
 Ile Ile Thr Ala Ile Arg Ser Leu Arg Glu Arg Gly Val Glu Phe Leu
 275 280 285
 Ala Val Pro Phe Thr Tyr Tyr Lys Gln Leu Gln Glu Lys Leu Lys Ser
 290 295 300
 Ala Lys Ile Arg Val Lys Glu Ser Ile Asp Val Leu Glu Glu Leu Lys
 305 310 315 320
 Ile Leu Val Asp Tyr Asp Glu Lys Gly Tyr Leu Leu Gln Ile Phe Thr
 325 330 335
 Lys Pro Met Gln Asp Arg Pro Thr Val Phe Leu Glu Val Ile Gln Arg
 340 345 350
 Asn Asn His Gln Gly Phe Gly Ala Gly Asn Phe Asn Ser Leu Phe Lys
 355 360 365
 Ala Phe Glu Glu Glu Gln Glu Leu Arg Gly Asn Leu Thr Asp Thr Asp
 370 375 380
 Pro Asn Gly Val Pro Phe Arg Leu
 385 390

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 392 amino acids
 (B) TYPE: amino acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```

Thr Thr Tyr Asn Asn Lys Gly Pro Lys Pro Glu Arg Gly Arg Phe Leu
1          5          10          15

His Phe His Ser Val Thr Phe Trp Val Gly Asn Ala Lys Gln Ala Ala
20          25          30

Ser Phe Tyr Cys Asn Lys Met Gly Phe Glu Pro Leu Ala Tyr Arg Gly
35          40          45

Leu Glu Thr Gly Ser Arg Glu Val Val Ser His Val Ile Lys Arg Gly
50          55          60

Lys Ile Val Phe Val Leu Cys Ser Ala Leu Asn Pro Trp Asn Lys Glu
65          70          75          80

Met Gly Asp His Leu Val Lys His Gly Asp Gly Val Lys Asp Ile Ala
85          90          95

Phe Glu Val Glu Asp Cys Asp His Ile Val Gln Lys Ala Arg Glu Arg
100         105         110

Gly Ala Lys Ile Val Arg Glu Pro Trp Val Glu Gln Asp Lys Phe Gly
115         120         125

Lys Val Lys Phe Ala Val Leu Gln Thr Tyr Gly Asp Thr Thr His Thr
130         135         140

Leu Val Glu Lys Ile Asn Tyr Thr Gly Arg Phe Leu Pro Gly Phe Glu
145         150         155         160

Ala Pro Thr Tyr Lys Asp Thr Leu Leu Pro Lys Leu Pro Arg Cys Asn
165         170         175

Leu Glu Ile Ile Asp His Ile Val Gly Asn Gln Pro Asp Gln Glu Met
180         185         190

Gln Ser Ala Ser Glu Trp Tyr Leu Lys Asn Leu Gln Phe His Arg Phe
195         200         205

Trp Ser Val Asp Asp Thr Gln Val His Thr Glu Tyr Ser Ser Leu Arg
210         215         220

Ser Ile Val Val Thr Asn Tyr Glu Glu Ser Ile Lys Met Pro Ile Asn
225         230         235         240

Glu Pro Ala Pro Gly Arg Lys Lys Ser Gln Ile Gln Glu Tyr Val Asp
245         250         255

Tyr Asn Gly Gly Ala Gly Val Gln His Ile Ala Leu Lys Thr Glu Asp
260         265         270

Ile Ile Thr Ala Ile Arg His Leu Arg Glu Arg Gly Thr Glu Phe Leu
275         280         285

Ala Ala Pro Ser Ser Tyr Tyr Lys Leu Leu Arg Glu Asn Leu Lys Ser
290         295         300

Ala Lys Ile Gln Val Lys Glu Ser Met Asp Val Leu Glu Glu Leu His
305         310         315         320

```

Ile Leu Val Asp Tyr Asp Glu Lys Gly Tyr Leu Leu Gln Ile Phe Thr
 325 330 335
 Lys Pro Met Gln Asp Arg Pro Thr Leu Phe Leu Glu Val Ile Gln Arg
 340 345 350
 His Asn His Gln Gly Phe Gly Ala Gly Asn Phe Asn Ser Leu Phe Lys
 355 360 365
 Ala Phe Glu Glu Glu Gln Ala Leu Arg Gly Asn Leu Thr Asp Leu Glu
 370 375 380
 Pro Asn Gly Val Arg Ser Gly Met
 385 390

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 376 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Tyr Trp Asp Lys Gly Pro Lys Pro Glu Arg Gly Arg Phe Leu His Phe
 1 5 10 15
 His Ser Val Thr Phe Trp Val Gly Asn Ala Lys Gln Ala Ala Ser Phe
 20 25 30
 Tyr Cys Asn Lys Met Gly Phe Glu Pro Leu Ala Tyr Lys Gly Leu Glu
 35 40 45
 Thr Gly Ser Arg Glu Val Val Ser His Val Ile Lys Gln Gly Lys Ile
 50 55 60
 Val Phe Val Leu Cys Ser Ala Leu Asn Pro Trp Asn Lys Glu Met Gly
 65 70 75 80
 Asp His Leu Val Lys His Gly Asp Gly Val Lys Asp Ile Ala Phe Glu
 85 90 95
 Val Glu Asp Cys Glu His Ile Val Gln Lys Ala Arg Glu Arg Gly Ala
 100 105 110
 Lys Ile Val Arg Glu Pro Trp Val Glu Glu Asp Lys Phe Gly Lys Val
 115 120 125
 Lys Phe Ala Val Leu Gln Thr Tyr Gly Asp Thr Thr His Thr Leu Val
 130 135 140
 Glu Lys Ile Asn Tyr Thr Gly Arg Phe Leu Pro Gly Phe Glu Ala Pro
 145 150 155 160
 Thr Tyr Lys Asp Thr Leu Leu Pro Lys Leu Pro Ser Cys Asn Leu Glu
 165 170 175
 Ile Ile Asp His Ile Val Gly Asn Gln Pro Asp Gln Glu Met Glu Ser
 180 185 190
 Ala Ser Glu Trp Tyr Leu Lys Asn Leu Gln Phe His Arg Phe Trp Ser
 195 200 205

Val Asp Asp Thr Gln Val His Thr Glu Tyr Ser Ser Leu Arg Ser Ile
 210 215 220
 Val Val Ala Asn Tyr Glu Glu Ser Ile Lys Met Pro Ile Asn Glu Pro
 225 230 235 240
 Ala Pro Gly Arg Lys Lys Ser Gln Ile Gln Glu Tyr Val Asp Tyr Asn
 245 250 255
 Gly Gly Ala Gly Val Gln His Ile Ala Leu Arg Thr Glu Asp Ile Ile
 260 265 270
 Thr Thr Ile Arg His Leu Arg Glu Arg Gly Met Glu Phe Leu Ala Val
 275 280 285
 Pro Ser Ser Tyr Tyr Arg Leu Leu Arg Glu Asn Leu Lys Thr Ser Lys
 290 295 300
 Ile Gln Val Lys Glu Asn Met Asp Val Leu Glu Glu Leu Lys Ile Leu
 305 310 315 320
 Val Asp Tyr Asp Glu Lys Gly Tyr Leu Leu Gln Ile Phe Thr Lys Pro
 325 330 335
 Met Gln Asp Arg Pro Thr Leu Phe Leu Glu Val Ile Gln Arg His Asn
 340 345 350
 His Gln Gly Phe Gly Ala Gly Asn Phe Asn Ser Leu Phe Lys Ala Phe
 355 360 365
 Glu Glu Glu Gln Ala Leu Arg Gly
 370 375

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1766 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Zea mays
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 261..1595
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

ACTAGTGTGTG AGAGCCTTCT GCGTTGGCAA TTGGCAGTAC AAGACAAATC ACATCCGCAA 60
 CCGCAACCAC AGAATCGTCC GTCCACGTGG CCCCATCAC TTCCCTTTAT TTACAGTCG 120
 TCCCCATCC CCAGGGCCAC CCACCAACAA GTGCAGTCAC CCGAGCCGCA AACTGCAGCT 180
 CTGCAAGCTA CAGAGGCCAC CACGAGTCCA CGACGCCACG CCCTCCGAGA GAAAGAGAAA 240

| | |
|---|------|
| GAGAAAACCA AAGCACGATA ATG CCC CCG ACC CCC ACA GCC GCC GCA GCC | 290 |
| Met Pro Pro Thr Pro Thr Ala Ala Ala Ala | |
| 1 5 10 | |
| GGC GCC GCC GTG GCG GCG GCA TCA GCA GCG GAG CAA GCG GCG TTC CGC | 338 |
| Gly Ala Ala Val Ala Ala Ser Ala Ala Glu Gln Ala Ala Phe Arg | |
| 15 20 25 | |
| CTC GTG GGC CAC CGC AAC TTC GTC CGC TTC AAC CCG CGC TCC GAC CGC | 386 |
| Leu Val Gly His Arg Asn Phe Val Arg Phe Asn Pro Arg Ser Asp Arg | |
| 30 35 40 | |
| TTC CAC ACG CTC GCG TTC CAC CAC GTG GAG CTC TGG TGC GCC GAC GCG | 434 |
| Phe His Thr Leu Ala Phe His His Val Glu Leu Trp Cys Ala Asp Ala | |
| 45 50 55 | |
| GCC TCC GCC GCG GGC CGC TTC TCC TTC GGC CTG GGC GCG CCG CTC GCC | 482 |
| Ala Ser Ala Ala Gly Arg Phe Ser Phe Gly Leu Gly Ala Pro Leu Ala | |
| 60 65 70 | |
| GCA CGC TCC GAC CTC TCC ACG GGC AAC TCC GCG CAC GCG TCC CTC CTG | 530 |
| Ala Arg Ser Asp Leu Ser Thr Gly Asn Ser Ala His Ala Ser Leu Leu | |
| 75 80 85 90 | |
| CTC CGC TCC GGC TCC CTC TCC TTC CTC TTC ACG GCG CCG TAC GCG CAC | 578 |
| Leu Arg Ser Gly Ser Leu Ser Phe Leu Phe Thr Ala Pro Tyr Ala His | |
| 95 100 105 | |
| GGC GCC GAC GCT GCC ACC GCC GCG CTG CCC TCC TTC TCC GCC GCC GCC | 626 |
| Gly Ala Asp Ala Ala Thr Ala Ala Leu Pro Ser Phe Ser Ala Ala Ala | |
| 110 115 120 | |
| GCG CGG CGC TTC GCA GCC GAC CAC GGC CTC GCG GTG CGC GCC GTC GCG | 674 |
| Ala Arg Arg Phe Ala Ala Asp His Gly Leu Ala Val Arg Ala Val Ala | |
| 125 130 135 | |
| CTC CGC GTC GCC GAC GCC GAG GAC GCC TTC CGC GCC AGC GTC GCG GCC | 722 |
| Leu Arg Val Ala Asp Ala Glu Asp Ala Phe Arg Ala Ser Val Ala Ala | |
| 140 145 150 | |
| GGG GCG CGC CCG GCG TTC GGC CCC GTC GAC CTC GGC CGC GGC TTC CGC | 770 |
| Gly Ala Arg Pro Ala Phe Gly Pro Val Asp Leu Gly Arg Gly Phe Arg | |
| 155 160 165 170 | |
| CTC GCC GAG GTC GAG CTC TAC GGC GAC GTC GTG CTC CGG TAC GTG AGC | 818 |
| Leu Ala Glu Val Glu Leu Tyr Gly Asp Val Val Leu Arg Tyr Val Ser | |
| 175 180 185 | |
| TAC CCG GAC GGC GCC GCG GGC GAG CCC TTC CTG CCG GGG TTC GAG GGC | 866 |
| Tyr Pro Asp Gly Ala Ala Gly Glu Pro Phe Leu Pro Gly Phe Glu Gly | |
| 190 195 200 | |
| GTG GCC AGC CCC GGG GCG GCC GAC TAC GGG CTG AGC AGG TTC GAC CAC | 914 |
| Val Ala Ser Pro Gly Ala Ala Asp Tyr Gly Leu Ser Arg Phe Asp His | |
| 205 210 215 | |
| ATC GTC GGC AAC GTG CCG GAG CTG GCG CCC GCC GCC TAC TTC GCC | 962 |
| Ile Val Gly Asn Val Pro Glu Leu Ala Pro Ala Ala Tyr Phe Ala | |
| 220 225 230 | |
| GGC TTC ACG GGG TTC CAC GAG TTC GCC GAG TTC ACG ACG GAG GAC GTG | 1010 |
| Gly Phe Thr Gly Phe His Glu Phe Ala Glu Phe Thr Thr Glu Asp Val | |
| 235 240 245 250 | |

GGC ACC GCG GAG AGC GGC CTC AAC TCC ATG GTG CTC GCC AAC AAC TCG 1058
 Gly Thr Ala Glu Ser Gly Leu Asn Ser Met Val Leu Ala Asn Ser 265
 255

GAG AAC GTG CTG CTC CCG CTC AAC GAG CCG GTG CAC GGC ACC AAG CGC 1106
 Glu Asn Val Leu Leu Pro Leu Asn Glu Pro Val His Gly Thr Lys Arg 280
 270

CGC AGC CAG ATA CAA ACG TTC CTG GAC CAC CAC GGC GGC CCC GGC GTG 1154
 Arg Ser Gln Ile Gln Thr Phe Leu Asp His His Gly Gly Pro Gly Val 295
 285

CAG CAC ATG GCG CTG GCC AGC GAC GAC GTG CTC AGG ACG CTG AGG GAG 1202
 Gln His Met Ala Leu Ala Ser Asp Asp Val Leu Arg Thr Leu Arg Glu 310
 300

ATG CAG GCG CCG TCG GCC ATG GCG GCC TTC GAG TTC ATG GCG CCT CCC 1250
 Met Gln Ala Arg Ser Ala Met Gly Gly Phe Glu Phe Met Ala Pro Pro 330
 315

ACA TCC GAC TAC TAT GAC GGC GTG AGG CCG CCG GCC GGG GAC GTG CTC 1298
 Thr Ser Asp Tyr Tyr Asp Gly Val Arg Arg Arg Ala Gly Asp Val Leu 345
 335

ACG GAA GCA CAG ATT AAG GAG TGC CAG GAG CTA GGG GTG CTG GTG GAC 1346
 Thr Glu Ala Gln Ile Lys Glu Cys Gln Glu Leu Gly Val Leu Val Asp 360
 350

AGG GAT GAC CAG GGC GTG CTG CTC CAA ATC TTC ACC AAG CCA GTG GGC 1394
 Arg Asp Asp Gln Gly Val Leu Leu Gln Ile Phe Thr Lys Pro Val Gly 375
 365

GAC AGG CCA ACG CTG TTC TTG GAA ATC ATC CAA AGG ATC GGG TGC ATG 1442
 Asp Arg Pro Thr Leu Phe Leu Glu Ile Ile Gln Arg Ile Gly Cys Met 390
 380

GAG AAG GAT GAG AAG GGG CAA GAA TAC CAA AAG GGT GGC TGC GGC GGC 1490
 Glu Lys Asp Glu Lys Gly Gln Glu Tyr Gln Lys Gly Gly Cys Gly Gly 410
 395

TTC GGC AAG GGA AAC TTC TCG CAG CTG TTC AAG TCC ATC GAG GAT TAT 1538
 Phe Gly Lys Gly Asn Phe Ser Gln Leu Phe Lys Ser Ile Glu Asp Tyr 425
 415

GAG AAG TCC CTT GAA GCC AAG CAA GCT GCT GCA GCA GCT GCA GCT CAG 1586
 Glu Lys Ser Leu Glu Ala Lys Gln Ala Ala Ala Ala Ala Ala Ala Gln 440
 430

GGA TCC TAG GACAGTGCTT GGAGACGAGC AACTGCTGTG GCACTTTGTA 1635
 Gly Ser

TCATGGAACA GAAATAATGA AGCGTGTCT TTGTGACACT TGACATGCAA ATGTTTGTGT 1695

TCTGTAACCG TTGAATATAT GCGACGATGC TATGATGGTG TAATAGATGG TAGAGAGGGT 1755

ACAACCCTGA T 1766

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 445 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Pro Pro Thr Pro Thr Ala Ala Ala Ala Gly Ala Ala Val Ala Ala
 1 5 10 15
 Ala Ser Ala Ala Glu Gln Ala Ala Phe Arg Leu Val Gly His Arg Asn
 20 25 30
 Phe Val Arg Phe Asn Pro Arg Ser Asp Arg Phe His Thr Leu Ala Phe
 35 40 45
 His His Val Glu Leu Trp Cys Ala Asp Ala Ala Ser Ala Ala Gly Arg
 50 55 60
 Phe Ser Phe Gly Leu Gly Ala Pro Leu Ala Ala Arg Ser Asp Leu Ser
 65 70 75 80
 Thr Gly Asn Ser Ala His Ala Ser Leu Leu Leu Arg Ser Gly Ser Leu
 85 90 95
 *Ser Phe Leu Phe Thr Ala Pro Tyr Ala His Gly Ala Asp Ala Ala Thr
 100 105 110
 Ala Ala Leu Pro Ser Phe Ser Ala Ala Ala Ala Arg Arg Phe Ala Ala
 115 120 125
 Asp His Gly Leu Ala Val Arg Ala Val Ala Leu Arg Val Ala Asp Ala
 130 135 140
 Glu Asp Ala Phe Arg Ala Ser Val Ala Ala Gly Ala Arg Pro Ala Phe
 145 150 155 160
 Gly Pro Val Asp Leu Gly Arg Gly Phe Arg Leu Ala Glu Val Glu Leu
 165 170 175
 Tyr Gly Asp Val Val Leu Arg Tyr Val Ser Tyr Pro Asp Gly Ala Ala
 180 185 190
 Gly Glu Pro Phe Leu Pro Gly Phe Glu Gly Val Ala Ser Pro Gly Ala
 195 200 205
 Ala Asp Tyr Gly Leu Ser Arg Phe Asp His Ile Val Gly Asn Val Pro
 210 215 220
 Glu Leu Ala Pro Ala Ala Tyr Phe Ala Gly Phe Thr Gly Phe His
 225 230 235 240
 Glu Phe Ala Glu Phe Thr Thr Glu Asp Val Gly Thr Ala Glu Ser Gly
 245 250 255
 Leu Asn Ser Met Val Leu Ala Asn Asn Ser Glu Asn Val Leu Leu Pro
 260 265 270
 Leu Asn Glu Pro Val His Gly Thr Lys Arg Arg Ser Gln Ile Gln Thr
 275 280 285
 Phe Leu Asp His His Gly Gly Pro Gly Val Gln His Met Ala Leu Ala
 290 295 300
 Ser Asp Asp Val Leu Arg Thr Leu Arg Glu Met Gln Ala Arg Ser Ala
 305 310 315 320
 Met Gly Gly Phe Glu Phe Met Ala Pro Pro Thr Ser Asp Tyr Tyr Asp
 325 330 335

Gly Val Arg Arg Arg Ala Gly Asp Val Leu Thr Glu Ala Gln Ile Lys
 340 345 350

Glu Cys Gln Glu Leu Gly Val Leu Val Asp Arg Asp Asp Gln Gly Val
 355 360 365

Leu Leu Gln Ile Phe Thr Lys Pro Val Gly Asp Arg Pro Thr Leu Phe
 370 375 380

Leu Glu Ile Ile Gln Arg Ile Gly Cys Met Glu Lys Asp Glu Lys Gly
 385 390 395 400

Gln Glu Tyr Gln Lys Gly Gly Cys Gly Gly Phe Gly Lys Gly Asn Phe
 405 410 415

Ser Gln Leu Phe Lys Ser Ile Glu Asp Tyr Glu Lys Ser Leu Glu Ala
 420 425 430

Lys Gln Ala Ala Ala Ala Ala Ala Ala Gln Gly Ser
 435 440

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1356 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Arabidopsis thaliana

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1254

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1..3
- (D) OTHER INFORMATION: /standard_name=
"translation initiation
codon"

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: 1252..1254
- (D) OTHER INFORMATION: /standard_name=
"translation termination
codon"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

| | |
|---|----|
| ATG TCC AAG TTC GTA AGA AAG AAT CCA AAG TCT GAT AAA TTC AAG GTT | 48 |
| Met Ser Lys Phe Val Arg Lys Asn Pro Lys Ser Asp Lys Phe Lys Val | |
| 1 5 10 15 | |
| AAG CGC TTC CAT CAC ATC GAG TTC TGG TGC GGC GAC GCA ACC AAC GTC | 96 |
| Lys Arg Phe His His Ile Glu Phe Trp Cys Gly Asp Ala Thr Asn Val | |
| 20 25 30 | |

| | |
|---|-----|
| GCT CGT CGC TTC TCC TGG GGT CTG GGG ATG AGA TTC TCC GCC AAA TCC Ala Arg Arg Phe Ser Trp Gly Leu Gly Met Arg Phe Ser Ala Lys Ser 35 40 45 | 144 |
| GAT CTT TCC ACC GGA AAC ATG GTT CAC GCC TCT TAC CTA CTC ACC TCC Asp Leu Ser Thr Gly Asn Met Val His Ala Ser Tyr Leu Leu Thr Ser 50 55 60 | 192 |
| GGT GAC CTC CGA TTC CTT TTC ACT GCT CCT TAC TCT CCG TCT CTC TCC Gly Asp Leu Arg Phe Leu Phe Thr Ala Pro Tyr Ser Pro Ser Leu Ser 65 70 75 80 | 240 |
| GCC GGA GAG ATT AAA CCG ACA ACC ACA GCT TCT ATC CCA AGT TTC GAT Ala Gly Glu Ile Lys Pro Thr Thr Thr Ala Ser Ile Pro Ser Phe Asp 85 90 95 | 288 |
| CAC GGC TCT TGT CGT TCC TTC TTC TCT TCA CAT GGT CTC GGT GTT AGA His Gly Ser Cys Arg Ser Phe Phe Ser Ser His Gly Leu Gly Val Arg 100 105 110 | 336 |
| GCC GTT GCG ATT GAA GTA GAA GAC-GCA GAG TCA GCT TTC TCC ATC AGT Ala Val Ala Ile Glu Val Glu Asp Ala Glu Ser Ala Phe Ser Ile Ser 115 120 125 | 384 |
| GTA GGT AAT GGC GCT ATT CCT TCG TCG GGT GGT ATC GTC CTC AAT GAA Val Ala Asn Gly Ala Ile Pro Ser Ser Pro Pro Ile Val Leu Asn Glu 130 135 140 | 432 |
| GCA GTT ACG ATC GCT GAG GTT AAA CTA TAC GGC GAT GTT GTT CTC CGA Ala Val Thr Ile Ala Glu Val Lys Leu Tyr Gly Asp Val Val Leu Arg 145 150 155 160 | 480 |
| TAT GTT AGT TAC AAA GCA GAA GAT ACC GAA AAA TCC GAA TTC TTG CCA Tyr Val Ser Tyr Lys Ala Glu Asp Thr Glu Lys Ser Glu Phe Leu Pro 165 170 175 | 528 |
| GGG TTC GAG CGT GTA GAG GAT GCG TCG TCG TTC CCA TTG GAT TAT GGT Gly Phe Glu Arg Val Glu Asp Ala Ser Ser Phe Pro Leu Asp Tyr Gly 180 185 190 | 576 |
| ATC CGG CGG CTT GAC CAC GCC GTG GGA AAC GTT CCT GAG CTT GGT CCG Ile Arg Arg Leu Asp His Ala Val Gly Asn Val Pro Glu Leu Gly Pro 195 200 205 | 624 |
| GCT TTA ACT TAT GTA GCG GGG TTC ACT GGT TTT CAC CAA TTC GCA GAG Ala Leu Thr Tyr Val Ala Glu Phe Thr Gly Phe His Gln Phe Ala Glu 210 215 220 | 672 |
| TTC ACA GCA GAC GAC GTT GGA ACC GCC GAG AGC GGT TTA AAT TCA GCG Phe Thr Ala Asp Asp Val Gly Thr Ala Glu Ser Gly Leu Asn Ser Ala 225 230 235 240 | 720 |
| GTC CTG GCT AGC AAT GAT GAA ATG GTT CTT CTA CCG ATT AAC GAG CCA Val Leu Ala Ser Asn Asp Glu Met Val Leu Leu Pro Ile Asn Glu Pro 245 250 255 | 768 |
| GTG CAC GGA ACA AAG AGG AAG AGT CAG ATT CAG ACG TAT TTG GAA CAT Val His Gly Thr Lys Arg Lys Ser Gln Ile Gln Thr Tyr Leu Glu His 260 265 270 | 816 |
| AAC GAA GGC GCA GGG CTA CAA CAT CTG GCT CTG ATG AGT GAA GAC ATA Asn Glu Gly Ala Gly Leu Gln His Leu Ala Leu Met Ser Glu Asp Ile 275 280 285 | 864 |

TTC AGG ACC CTG AGA GAG ATG AGG AAG AGG AGC AGT ATT GGA GGA TTC 912
 Phe Arg Thr Leu Arg Glu Met Arg Lys Arg Ser Ser Ile Gly Gly Phe
 290 295 300

GAC TTC ATG CCT TCT CCT CCG CCT ACT TAC TAC CAG AAT CTC AAG AAA 960
 Asp Phe Met Pro Ser Pro Pro Pro Thr Tyr Gln Asn Leu Lys Lys
 305 310 315 320

CGG GTC GGC GAC GTG CTC AGC GAT GAT CAG ATC AAG GAG TGT GAG GAA 1008
 Arg Val Gly Asp Val Leu Ser Asp Asp Gln Ile Lys Glu Cys Glu Glu
 325 330 335

TTA GGG ATT CTT GTA GAC AGA GAT GAT CAA GGG ACG TTG CTT CAA ATC 1056
 Leu Gly Ile Leu Val Asp Arg Asp Asp Gln Gly Thr Leu Leu Gln Ile
 340 345 350

TTC ACA AAA CCA CTA GGT GAC AGG CCG ACG ATA TTT ATA GAG ATA ATC 1104
 Phe Thr Lys Pro Leu Gly Asp Arg Pro Thr Ile Phe Ile Glu Ile Ile
 355 360 365

CAG AGA GTG GGA TGC ATG ATG AAA GAT GAG-GAA GGG AAG GCT TAC CAG 1152
 Gln Arg Val Gly Cys Met Met Lys Asp Gln Glu Gly Lys Ala Tyr Gln
 370 375 380

AGT GGA GGA TGT GGT GGT TTT GGC AAA GGC AAT TTC TCT GAG CTC TTC 1200
 Ser Gly Gly Cys Gly Gly Phe Gly Lys Gly Asn Phe Ser Glu Leu Phe
 385 390 395 400

AAG TCC ATT GAA GAA TAC GAA AAG ACT CTT GAA GCC AAA CAG TTA GTG 1248
 Lys Ser Ile Glu Glu Tyr Glu Lys Thr Leu Glu Ala Lys Gln Leu Val
 405 410 415

GGA TGA ACAAGAAGAA GAACCAACTA AAGGATTGTG TAATTAATGT AAAACTGTTT 1304
 Gly *

TATCTTATCA AAACAATGTA TACAACATCT CATTAAAAA CGAGATCAAT CC 1356

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 418 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Met Ser Lys Phe Val Arg Lys Asn Pro Lys Ser Asp Lys Phe Lys Val
 1 5 10 15

Lys Arg Phe His His Ile Glu Phe Trp Cys Gly Asp Ala Thr Asn Val
 20 25 30

Ala Arg Arg Phe Ser Trp Gly Leu Gly Met Arg Phe Ser Ala Lys Ser
 35 40 45

Asp Leu Ser Thr Gly Asn Met Val His Ala Ser Tyr Leu Leu Thr Ser
 50 55 60

Gly Asp Leu Arg Phe Leu Phe Thr Ala Pro Tyr Ser Pro Ser Leu Ser
 65 70 75 80

Ala Gly Glu Ile Lys Pro Thr Thr Thr Ala Ser Ile Pro Ser Phe Asp
 85 90 95

His Gly Ser Cys Arg Ser Phe Phe Ser Ser His Gly Leu Gly Val Arg
 100 105 110
 Ala Val Ala Ile Glu Val Glu Asp Ala Glu Ser Ala Phe Ser Ile Ser
 115 120 125
 Val Ala Asn Gly Ala Ile Pro Ser Ser Pro Pro Ile Val Leu Asn Glu
 130 135 140
 Ala Val Thr Ile Ala Glu Val Lys Leu Tyr Gly Asp Val Val Leu Arg
 145 150 155 160
 Tyr Val Ser Tyr Lys Ala Glu Asp Thr Glu Lys Ser Glu Phe Leu Pro
 165 170 175
 Gly Phe Glu Arg Val Glu Asp Ala Ser Ser Phe Pro Leu Asp Tyr Gly
 180 185 190
 Ile Arg Arg Leu Asp His Ala Val Gly Asn Val Pro Glu Leu Gly Pro
 195 200 205
 Ala Leu Thr Tyr Val Ala Gly Phe Thr Gly Phe His Gln Phe Ala Glu
 210 215 220
 Phe Thr Ala Asp Asp Val Gly Thr Ala Glu Ser Gly Leu Asn Ser Ala
 225 230 235 240
 Val Leu Ala Ser Asn Asp Glu Met Val Leu Leu Pro Ile Asn Glu Pro
 245 250 255
 Val His Gly Thr Lys Arg Lys Ser Gln Ile Gln Thr Tyr Leu Glu His
 260 265 270
 Asn Glu Gly Ala Gly Leu Gln His Leu Ala Leu Met Ser Glu Asp Ile
 275 280 285
 Phe Arg Thr Leu Arg Glu Met Arg Lys Arg Ser Ser Ile Gly Gly Phe
 290 295 300
 Asp Phe Met Pro Ser Pro Pro Pro Thr Tyr Tyr Gln Asn Leu Lys Lys
 305 310 315 320
 Arg Val Gly Asp Val Leu Ser Asp Asp Gln Ile Lys Glu Cys Glu Glu
 325 330 335
 Leu Gly Ile Leu Val Asp Arg Asp Asp Gln Gly Thr Leu Leu Gln Ile
 340 345 350
 Phe Thr Lys Pro Leu Gly Asp Arg Pro Thr Ile Phe Ile Glu Ile Ile
 355 360 365
 Gln Arg Val Gly Cys Met Met Lys Asp Glu Glu Gly Lys Ala Tyr Gln
 370 375 380
 Ser Gly Gly Cys Gly Gly Phe Gly Lys Gly Asn Phe Ser Glu Leu Phe
 385 390 395 400
 Lys Ser Ile Glu Glu Tyr Glu Lys Thr Leu Glu Ala Lys Gln Leu Val
 405 410 415
 Gly *

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1448 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Arabidopsis thaliana
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 9..1346
- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 9..11
 - (D) OTHER INFORMATION: /standard_name="translation initiation codon"
- (ix) FEATURE:
 - (A) NAME/KEY: misc_feature
 - (B) LOCATION: 1344..1346
 - (D) OTHER INFORMATION: /standard_name="translation termination codon"
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```

TGAAATCA ATG GGC CAC CAA AAC GCC GCC GTT TCA GAG AAT CAA AAC CAT      50
  Met Gly His Gln Asn Ala Ala Val Ser Glu Asn Gln Asn His
    1             5             10

GAT GAC GGC GCT GCG TCG TCG CCG GGA TTC AAG CTC GTC GGA TTT TCC      98
Asp Asp Gly Ala Ala Ser Ser Pro Gly Phe Lys Leu Val Gly Phe Ser
  15             20             25             30

AAG TTC GTA AGA AAG AAT CCA AAG TCT GAT AAA TTC AAG GTT AAG CGC      146
Lys Phe Val Arg Lys Asn Pro Lys Ser Asp Lys Phe Lys Val Lys Arg
          35             40             45

TTC CAT CAC ATC GAG TTC TGG TGC GGC GAC GCA ACC AAC GTC GCT CGT      194
Phe His His Ile Glu Phe Trp Cys Gly Asp Ala Thr Asn Val Ala Arg
          50             55             60

CGC TTC TCC TGG GGT CTG GGG ATG AGA TTC TCC GCC AAA TCC GAT CTT      242
Arg Phe Ser Trp Gly Leu Gly Met Arg Phe Ser Ala Lys Ser Asp Leu
          65             70             75

TCC ACC GGA AAC ATG GTT CAC GCC TCT TAC CTA CTC ACC TCC GGT GAC      290
Ser Thr Gly Asn Met Val His Ala Ser Tyr Leu Leu Thr Ser Gly Asp
          80             85             90

CTC CGA TTC CTT TTC ACT GCT CCT TAC TCT CCG TCT CTC TCC GCC GGA      338
Leu Arg Phe Leu Phe Thr Ala Pro Tyr Ser Pro Ser Leu Ser Ala Gly
          95             100             105             110

GAG ATT AAA CCG ACA ACC ACA GCT TCT ATC CCA AGT TTC GAT CAC GGC      386
Glu Ile Lys Pro Thr Thr Thr Ala Ser Ile Pro Ser Phe Asp His Gly
          115             120             125

```

| | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| TCT | TGT | CGT | TCC | TTC | TTC | TCT | TCA | CAT | GGT | CTC | GGT | GTT | AGA | GCC | GTT | 434 |
| Ser | Cys | Arg | Ser | Phe | Phe | Ser | Ser | His | Gly | Leu | Gly | Val | Arg | Ala | Val | |
| | | | 130 | | | | | 135 | | | | | 140 | | | |
| GCG | ATT | GAA | GTA | GAA | GAC | GCA | GAG | TCA | GCT | TTC | TCC | ATC | AGT | GTA | GCT | 482 |
| Ala | Ile | Glu | Val | Glu | Asp | Ala | Glu | Ser | Ala | Phe | Ser | Ile | Ser | Val | Ala | |
| | | 145 | | | | | 150 | | | | | 155 | | | | |
| AAT | GGC | GCT | ATT | CCT | TCG | TCG | CCT | CCT | ATC | GTC | CTC | AAT | GAA | GCA | GTT | 530 |
| Asn | Gly | Ala | Ile | Pro | Ser | Ser | Pro | Pro | Ile | Val | Leu | Asn | Glu | Ala | Val | |
| | 160 | | | | | | 165 | | | | 170 | | | | | |
| ACG | ATC | GCT | GAG | GTT | AAA | CTA | TAC | GGC | GAT | GTT | GTT | CTC | CGA | TAT | GTT | 578 |
| Thr | Ile | Ala | Glu | Val | Lys | Leu | Tyr | Gly | Asp | Val | Val | Leu | Arg | Tyr | Val | |
| | 175 | | | | | 180 | | | | 185 | | | | | 190 | |
| AGT | TAC | AAA | GCA | GAA | GAT | ACC | GAA | AAA | TCC | GAA | TTC | TTG | CCA | GGG | TTC | 626 |
| Ser | Tyr | Lys | Ala | Glu | Asp | Thr | Glu | Lys | Ser | Glu | Phe | Leu | Pro | Gly | Phe | |
| | | | | 195 | | | | | 200 | | | | | 205 | | |
| GAG | CGT | GTA | GAG | GAT | GCG | TCG | TCG | TTC | CCA | TTG | GAT | TAT | GGT | ATC | CGG | 674 |
| Glu | Arg | Val | Glu | Asp | Ala | Ser | Ser | Phe | Pro | Leu | Asp | Tyr | Gly | Ile | Arg | |
| | | | 210 | | | | | 215 | | | | | 220 | | | |
| CGG | CTT | GAC | CAC | GCC | GTG | GGA | AAC | GTT | CCT | GAG | CTT | GGT | CCG | GCT | TTA | 722 |
| Arg | Leu | Asp | His | Ala | Val | Gly | Asn | Val | Pro | Glu | Leu | Gly | Pro | Ala | Leu | |
| | | 225 | | | | | 230 | | | | | 235 | | | | |
| ACT | TAT | GTA | GCG | GGG | TTC | ACT | GGT | TTT | CAC | CAA | TTC | GCA | GAG | TTC | ACA | 770 |
| Thr | Tyr | Val | Ala | Gly | Phe | Thr | Gly | Phe | His | Gln | Phe | Ala | Glu | Phe | Thr | |
| | 240 | | | | | 245 | | | | | 250 | | | | | |
| GCA | GAC | GAC | GTT | GGA | ACC | GCC | GAG | AGC | GGT | TTA | AAT | TCA | GCG | GTC | CTG | 818 |
| Ala | Asp | Asp | Val | Gly | Thr | Ala | Glu | Ser | Gly | Leu | Asn | Ser | Ala | Val | Leu | |
| | 255 | | | | 260 | | | | 265 | | | | | | 270 | |
| GCT | AGC | AAT | GAT | GAA | ATG | GTT | CTT | CTA | CCG | ATT | AAC | GAG | CCA | GTG | CAC | 866 |
| Ala | Ser | Asn | Asp | Glu | Met | Val | Leu | Leu | Pro | Ile | Asn | Glu | Pro | Val | His | |
| | | | | 275 | | | | | 280 | | | | | 285 | | |
| GGA | ACA | AAG | AGG | AAG | AGT | CAG | ATT | CAG | ACG | TAT | TTG | GAA | CAT | AAC | GAA | 914 |
| Gly | Thr | Lys | Arg | Lys | Ser | Gln | Ile | Gln | Thr | Tyr | Leu | Glu | His | Asn | Glu | |
| | | | 290 | | | | | 295 | | | | | 300 | | | |
| GGC | GCA | GGG | CTA | CAA | CAT | CTG | GCT | CTG | ATG | AGT | GAA | GAC | ATA | TTC | AGG | 962 |
| Gly | Ala | Gly | Leu | Gln | His | Leu | Ala | Leu | Met | Ser | Glu | Asp | Ile | Phe | Arg | |
| | | 305 | | | | | 310 | | | | | 315 | | | | |
| ACC | CTG | AGA | GAG | ATG | AGG | AAG | AGG | AGC | AGT | ATT | GGA | GGA | TTC | GAC | TTC | 1010 |
| Thr | Leu | Arg | Glu | Met | Arg | Lys | Arg | Ser | Ser | Ile | Gly | Gly | Phe | Asp | Phe | |
| | 320 | | | | | 325 | | | | | 330 | | | | | |
| ATG | CCT | TCT | CCT | CCG | CCT | ACT | TAC | TAC | CAG | AAT | CTC | AAG | AAA | CGG | GTC | 1058 |
| Met | Pro | Ser | Pro | Pro | Pro | Thr | Tyr | Tyr | Gln | Asn | Leu | Lys | Lys | Arg | Val | |
| | 335 | | | | 340 | | | | 345 | | | | | | 350 | |
| GGC | GAC | GTG | CTC | AGC | GAT | GAT | CAG | ATC | AAG | GAG | TGT | GAG | GAA | TTA | GGG | 1106 |
| Gly | Asp | Val | Leu | Ser | Asp | Asp | Gln | Ile | Lys | Glu | Cys | Glu | Glu | Leu | Gly | |
| | | | | 355 | | | | | 360 | | | | | 365 | | |
| ATT | CTT | GTA | GAC | AGA | GAT | GAT | CAA | GGG | ACG | TTG | CTT | CAA | ATC | TTC | ACA | 1154 |
| Ile | Leu | Val | Asp | Arg | Asp | Asp | Gln | Gly | Thr | Leu | Leu | Gln | Ile | Phe | Thr | |
| | | | 370 | | | | | 375 | | | | | 380 | | | |

AAA CCA CTA GGT GAC AGG CCG ACG ATA TTT ATA GAG ATA ATC CAG AGA 1202
 Lys Pro Leu Gly Asp Arg Pro Thr Ile Phe Ile Glu Ile Ile Gln Arg
 385 390 395
 GTA GGA TGC ATG ATG AAA GAT GAG GAA GGG AAG GCT TAC CAG AGT GGA 1250
 Val Gly Cys Met Met Lys Asp Glu Glu Gly Lys Ala Tyr Gln Ser Gly
 400 405 410
 GGA TGT GGT GGT TTT GGC AAA GGC AAT TTC TCT GAG CTC TTC AAG TCC 1298
 Gly Cys Gly Gly Phe Gly Lys Gly Asn Phe Ser Glu Leu Phe Lys Ser
 415 420 425 430
 ATT GAA GAA TAC GAA AAG ACT CTT GAA GCC AAA CAG TTA GTG GGA TGA 1346
 Ile Glu Glu Tyr Glu Lys Thr Leu Glu Ala Lys Gln Leu Val Gly *
 435 440 445
 ACAAGAAGAA GAACCAACTA AAGGATTGTG TAATTAATGT AAAACTGTTT TATCTTATCA 1406
 AAACAATGTA TACAACATCT CATTAAAAA CGAGATCAAT CC 1448

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 446 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Met Gly His Gln Asn Ala Ala Val Ser Glu Asn Gln Asn His Asp Asp
 1 5 10 15
 Gly Ala Ala Ser Ser Pro Gly Phe Lys Leu Val Gly Phe Ser Lys Phe
 20 25 30
 Val Arg Lys Asn Pro Lys Ser Asp Lys Phe Lys Val Lys Arg Phe His
 35 40 45
 His Ile Glu Phe Trp Cys Gly Asp Ala Thr Asn Val Ala Arg Arg Phe
 50 55 60
 Ser Trp Gly Leu Gly Met Arg Phe Ser Ala Lys Ser Asp Leu Ser Thr
 65 70 75 80
 Gly Asn Met Val His Ala Ser Tyr Leu Leu Thr Ser Gly Asp Leu Arg
 85 90 95
 Phe Leu Phe Thr Ala Pro Tyr Ser Pro Ser Leu Ser Ala Gly Glu Ile
 100 105 110
 Lys Pro Thr Thr Thr Ala Ser Ile Pro Ser Phe Asp His Gly Ser Cys
 115 120 125
 Arg Ser Phe Phe Ser Ser His Gly Leu Gly Val Arg Ala Val Ala Ile
 130 135 140
 Glu Val Glu Asp Ala Glu Ser Ala Phe Ser Ile Ser Val Ala Asn Gly
 145 150 155 160
 Ala Ile Pro Ser Ser Pro Pro Ile Val Leu Asn Glu Ala Val Thr Ile
 165 170 175
 Ala Glu Val Lys Leu Tyr Gly Asp Val Val Leu Arg Tyr Val Ser Tyr
 180 185 190

Lys Ala Glu Asp Thr Glu Lys Ser Glu Phe Leu Pro Gly Phe Glu Arg
 195 200 205
 Val Glu Asp Ala Ser Ser Phe Pro Leu Asp Tyr Gly Ile Arg Arg Leu
 210 215 220
 Asp His Ala Val Gly Asn Val Pro Glu Leu Gly Pro Ala Leu Thr Tyr
 225 230 235 240
 Val Ala Gly Phe Thr Gly Phe His Gln Phe Ala Glu Phe Thr Ala Asp
 245 250 255
 Asp Val Gly Thr Ala Glu Ser Gly Leu Asn Ser Ala Val Leu Ala Ser
 260 265 270
 Asn Asp Glu Met Val Leu Leu Pro Ile Asn Glu Pro Val His Gly Thr
 275 280 285
 Lys Arg Lys Ser Gln Ile Gln Thr Tyr Leu Glu His Asn Glu Gly Ala
 290 295 300
 Gly Leu Gln His Leu Ala Leu Met Ser Glu Asp Ile Phe Arg Thr Leu
 305 310 315 320
 Arg Glu Met Arg Lys Arg Ser Ser Ile Gly Gly Phe Asp Phe Met Pro
 325 330 335
 Ser Pro Pro Pro Thr Tyr Tyr Gln Asn Leu Lys Lys Arg Val Gly Asp
 340 345 350
 Val Leu Ser Asp Asp Gln Ile Lys Glu Cys Glu Glu Leu Gly Ile Leu
 355 360 365
 Val Asp Arg Asp Asp Gln Gly Thr Leu Leu Gln Ile Phe Thr Lys Pro
 370 375 380
 Leu Gly Asp Arg Pro Thr Ile Phe Ile Glu Ile Ile Gln Arg Val Gly
 385 390 395 400
 Cys Met Met Lys Asp Glu Glu Gly Lys Ala Tyr Gln Ser Gly Gly Cys
 405 410 415
 Gly Gly Phe Gly Lys Gly Asn Phe Ser Glu Leu Phe Lys Ser Ile Glu
 420 425 430
 Glu Tyr Glu Lys Thr Leu Glu Ala Lys Gln Leu Val Gly
 435 440 445

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 513 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: NO
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: *Vernonia galamensis*
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: vs1.pk0015.b2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

| | |
|---|-----|
| CCACACCGAT TGCCGGAAC TCACCGCCTC TCACGGCCTT GCAGTCCGAG CAATCGCCAT | 60 |
| TGAAGTCGAT GACGCCGAAT TAGCTTTCTC CGTCAGCGTC TCTACGGCG CTAAACCCTC | 120 |
| CGCTGCTCCT GTAACCCTTG GAAACAACGA CGTCGTATTG TCTGAAGTTA AGCTTTACGG | 180 |
| CGATGTCGCT TTCCGGTACA TAAGTTACAA AAATCCGAAC TATACATCTT CCTTTTGGC | 240 |
| CGGGTTGAG CCCGTTGAAA AGACGTCGTC GTTTTATGAC CTTGACTACG GTATCCGCCG | 300 |
| TTTGGACCAC GCCGTAGGNA ACGTCCCTGA GCTTGCTTCG GCAGTGGACT ACGTGAAATC | 360 |
| ATTCACCGGA TTCCATGAGT TCGCCGAATT CACCGCGGAG GACGTCGGA CGACCGAGAG | 420 |
| GGAAGTGAAT TCGGTCGTTT TAGCTTGCAA CAGTGAGATG GTCTTGATTC CGATGAACGA | 480 |
| GCCGGTGTAC GGAANAAAAG GAAGNAGCCA GAT | 513 |

CLAIMS

1. An isolated nucleic acid fragment encoding a plant *p*-hydroxy-phenylpyruvate dioxygenase enzyme, the fragment comprising a nucleotide sequence selected from the group consisting of
 - 5 nucleotide sequences encoding a polypeptide comprising the amino acid sequences set forth in SEQ ID NO:3, SEQ ID NO:11, SEQ ID NO:13, and SEQ ID NO:15 and
 - modified nucleotide sequences essentially similar to the nucleotide sequences of SEQ ID NO:2, SEQ ID NO 10, SEQ ID NO:12 and
 - 10 SEQ ID NO:14 containing deletions, insertions, or substitutions in the sequence that do not affect the functional properties of the encoded protein.
2. An isolated nucleic acid fragment encoding a plant *p*-hydroxyphenylpyruvate dioxygenase enzyme, the fragment comprising a nucleotide sequence as set forth in SEQ ID NO:14.
- 15 3. A chimeric gene comprising the nucleic acid fragment of Claims 1 or 2 operably linked to at least one suitable regulatory sequence.
4. The chimeric gene of Claim 3 wherein at least one suitable regulatory sequence directs gene expression in a microorganism.
- 20 5. The chimeric gene of Claim 3 wherein the at least one suitable regulatory sequence directs gene expression in a plant.
6. A plasmid vector comprising the nucleic acid fragment of Claims 1 or 2 operably linked to at least one suitable regulatory sequence.
7. A transformed host cell comprising a host cell and the plasmid vector
- 25 of Claim 6.
8. The transformed host cell of Claim 7 wherein the host cell is derived from a plant or is a microorganism.
9. The transformed host cell of Claim 8 wherein the microorganism is *E. coli*.
- 30 10. A transformed plant tolerant to contact with at least one compound that inhibits the rate of the reaction of *p*-hydroxyphenylpyruvate dioxygenase enzyme in a non-transformed plant, the transformed plant comprising the chimeric gene of Claim 3 and a host plant.
11. The transformed plant of Claim 10 wherein the host plant is a cereal
- 35 crop plant.
12. A method to identify a compound useful for its ability to inhibit the rate of the reaction of *p*-hydroxyphenylpyruvate dioxygenase enzyme comprising:
 - (a) transforming a host cell with the plasmid vector of Claim 6;

- (b) facilitating expression of the nucleic acid fragment encoding the plant *p*-hydroxyphenylpyruvate dioxygenase enzyme;
 - (c) contacting the expressed enzyme from step (b) with a test compound; and
 - 5 (d) evaluating the capacity of the test compound to inhibit the rate of the reaction of *p*-hydroxyphenylpyruvate dioxygenase enzyme.
13. The method of Claim 12 wherein evaluating the capacity of the test compound to inhibit the rate of the reaction of *p*-hydroxyphenylpyruvate dioxygenase enzyme is accomplished by measuring oxygen utilization, carbon
- 10 dioxide release, homogentisate production, loss of *p*-hydroxyphenylpyruvate or maleylacetoacetate production.
14. The method of Claim 12 wherein the transformed host cell is an *E. coli* that comprises a chimeric gene encoding a plant *p*-hydroxyphenylpyruvate dioxygenase enzyme.
- 15 15. A compound that inhibits the activity of a plant *p*-hydroxyphenylpyruvate dioxygenase enzyme, the compound identified by the method of Claim 14.
16. A method for imparting tolerance to a plant to at least one compound that inhibits the rate of reaction of *p*-hydroxyphenylpyruvate dioxygenase enzyme
- 20 comprising:
- (a) transforming a host plant cell with a chimeric gene comprising a nucleic acid fragment encoding plant *p*-hydroxyphenylpyruvate dioxygenase, and
 - (b) expressing the chimeric gene in an amount effective to render
 - 25 the transformed plant substantially tolerant to the at least one compound that inhibits the rate of reaction of *p*-hydroxyphenylpyruvate dioxygenase.
17. A method for the microbial production of active plant *p*-hydroxyphenylpyruvate dioxygenase enzyme comprising:
- 30 (a) stably transforming a microorganism with the chimeric gene of Claim 4 encoding the plant *p*-hydroxyphenylpyruvate dioxygenase;
- (b) facilitating expression by the chimeric gene for a suitable period; and
 - 35 (c) recovering active plant *p*-hydroxyphenylpyruvate dioxygenase enzyme.
18. A method to overexpress *p*-hydroxyphenylpyruvate dioxygenase enzyme in a plant comprising:

- 5 (a) stably transforming a host plant cell with a chimeric DNA molecule comprising at least one copy of a suitable promoter to drive expression of an associated coding sequence in a plant cell operably linked to at least one copy of a homologous or heterologous coding sequence encoding *p*-hydroxyphenyl-pyruvate dioxygenase; and

(b) growing the transformed host plant cell of step (a).

19. The method of Claim 18 wherein the chimeric DNA molecule is the chimeric gene of Claim 5.

10 20. An isolated nucleic acid fragment comprising a member selected from the group consisting of:

- 15 (a) an isolated nucleic acid fragment as set forth in SEQ ID NO:16;
(b) an isolated nucleic acid fragment that is essentially similar to an isolated nucleic acid fragment as set forth in SEQ ID NO:16;
and
(c) an isolated nucleic acid fragment that is complementary to (a) or (b).

1/6

FIG. 1

1 CAAGAAACGNGTCGNCGACGTGCTCAGCGATGATCAGATCAAGGAGTGTGAGGAATTAGG
61 GATTCTTNTAGACAGAGATGATCAAGGGACGTTNCTTCAAATCTNCACAAAACCACTAGG
121 TGACAGGCCGACGNTATTTATAGAGATAATCCAGAGNGTAGGATGCATGATGAAAGATGT
181 GGAAGGGANGGCTTACCAGAGTGGAGNATNTNGTGGTTTTGGCAAAGGCAATT

2/6

FIG. 2

1 TGAAATCAATGGGCCACCAAACGCCGCCGTTTCAGAGAATCAAAACCATGATGACGGCG
61 CTGCGTCGTCGCCGGGATTCAAGCTCGTCGGATTTTCCAAGTTCGTAAGAAAGAATCCAA
121 AGTCTGATAAATTCAAGGTTAAGCGCTTCCATCACATCGAGTTCGGTGCGGGGACGCAA
Eco47III
181 CCAACGTCGCTCGTCGCTTCTCCTGGGGTCTGGGGATGAGATTCTCCGCCAAATCCGATC
241 TTTCCACCGGAAACATGGTTACGCCTCTTACCTACTCACCTCCGGTGAACCTCCGATTCC
301 TTTTCACTGCTCCTTACTCTCCGTCTCTCTCCGGCGGAGAGATTAAACCGACAACCACAG
361 GTTCTATCCCAAGTTTCGATCACGGGTCTTGTCGGTCCTTCTTCTTTCACATGGTCTCG
421 GTGTTAGACCCGTTGCGATTGAAGTAGAAGACGCGGAGTCAAGCTTCTCCATCAGTGTAG
481 CTAATGGCGCTATTCTTCGTCGCTCCTATCGTCCTCAATGAAGCAGTTACGATCGCTG
541 AGGTTAACTATACGGGATGTTGTTCTCCGATATGTTAGTTACAAAGCAGAAGATACCG
601 AAAAATCCGAATTCTTGCCAGGGTTCGAGCGTGTAGAGGATGCGTCGTCGTTCCCATTTGG
EcoRI
661 ATTATGGTATCCGGCGGCTTGACCACGCCGTGGGAAACGTTCTGAGCTTGGTCCGGCTT
721 TAACTTATGTAGCGGGGTTCACTGGTTTTACCAATTTCGCAGAGTTCACAGCAGACGACG
781 TTGGAACCGCCGAGAGCGGTTTAAATTCAGCGGTCTGGCTAGCAATGATGAAATGGTTC
NheI
841 TTCTACCGATTAAAGAGCCAGTGCACGGAAACAAAGAGGAAGAGTCAGATTACAGCGTATT
901 TGGAACATAACGAAGGCGCAGGGCTACAACATCTGGCTCTGATGAGTGAAGACATATTCA
961 GGACCCTGAGAGAGATGAGGAAGAGGAGCAGTATTGGAGGATTGACTTCATGCCTTCTC
1021 CTCCGCCTACTTACTACCAGAATCTCAAGAAACGGGTCCGGCAGCTGCTCAGCGATGATC
1081 AGATCAAGGAGTGTGAGGAATTAGGGATTCTTGTTAGACAGAGATGATCAAGGGACGTTGC
1141 TTCAAATCTTCACAAAACCACTAGGTGACAGGCCGACGATATTTATAGAGATAATCCAGA
1201 GAGTAGGATGCATGATGAAAGATGAGGAAGGAAGGCTTACCAGAGTGGAGGATGTGGTG
1261 GTTTTGCCAAAGGCAATTTCTCTGAGCTCTTCAAGTCCATTGAAGAATACGAAAAGACTC
1321 TTGAAGCCAAACAGTTAGTGGGATGAACAAGAAGAAGAACCAACTAAAGGATTGTGTAAT
1381 TAATGTAAAACTGTTTTATCTTATCAAAACAATGTATACAACATCTCATTTAAAAACGAG
1441 ATCAATCC

3/6

FIG. 3A

| | | | | | | |
|-------------|------------|------------|-------------|--------------|------------|--------|
| | | | | | | 50 |
| Arabidopsis | MGHQNAAVS | ENQNHDDGAA | SSFGKLVGF | SKFVRKNPKS | DKEKVKRFHH | |
| Corn | MBPTPTAAAA | GAAVAAASAA | EQAAERLVGH | RNEVRENPRS | DRFHTLAFHH | |
| Rat | | | | YWOKGPKP | ERGRFLHFHS | |
| Mouse | | | | M TTYNNKGPKP | ERGRFLHFHS | |
| Human | | | | M TYSOKGAKP | ERGRFLHFHS | |
| Pig | | | | M TSYOKGKGP | ERGRFLHFHS | ** |
| | | | | | | 100 |
| Arabidopsis | IEFWCGDATN | VARRFWSWLG | MRFSAKSDLS | TGNMVHASYL | LTSGDLRFLF | |
| Corn | VELWCADAAS | AAGRFSEGLG | APLAARSDLS | TGNSAHASLL | LRSGSLSFLF | |
| Rat | VTFWVGNAKQ | AASFYCNKMG | FEPLAYKGLE | TGSREVVSHV | IKQKGIVFVL | |
| Mouse | VTFWVGNAKQ | AASFYCNKMG | FEPLAYRGLE | TGSREVVSHV | IKRGKIVFVL | |
| Human | VTFWVGNAKQ | AASFYCSKMG | FEPLAYRGLE | TGSREVVSHV | IKQKGIVFVL | |
| Pig | VTFWVGNAKQ | AASYYSKIG | FEPLAYKGLE | TGSREVVSHV | VKQDKIVFVE | |
| | | | | | | 150 |
| Arabidopsis | TAPYSPSLSA | GEIKPTTTAS | IPSFHDGSCR | SFFSSHGLGV | RAVAIEVEDA | |
| Corn | TAPYAHGADA |ATAA | LPSFSAAAAAR | RFAADHGLAV | RAVALRVADA | |
| Rat | CSALNPW... | |NKEMG | DHLVKHGDGV | KDIAFEVEDC | |
| Mouse | CSALNPW... | |NKEMG | DHLVKHGDGV | KDIAFEVEDC | |
| Human | SSALNPW... | |NKEMG | DHLVKHGDGV | KDIAFEVEDC | |
| Pig | SSALNPW... | |NKEMG | DHLVKHGDGV | KDIAFEVEDC | ** * * |
| | | | | | | 200 |
| Arabidopsis | ESAFSISVAN | GAIPSSPIV | LNEAVTIAEV | KLYGDVVLRY | VSYKAEDTEK | |
| Corn | EDAFRASVAA | GARPAFGPVD | LGRGFRLAEV | ELYGDVVLRY | VSY.PDGAAG | |
| Rat | EHIVQKARER | GAKIVREPWW | EEDKFGKVKE | AVLQTYGDTT | HTLVEKINYT | |
| Mouse | DHIVQKARER | GAKIVREPWW | EQDKFGKVKE | AVLQTYGDTT | HTLVEKINYT | |
| Human | DYIVQKARER | GAKINREPWW | EQDKFGKVKE | AVLQTYGDTT | HTLVEKMNYI | |
| Pig | DYIVQKARER | GAIIVREPWI | EQDKFGKVKE | AVLQTFGDTT | HTLVEKMNYT | ** * |
| | | | | | | 250 |
| Arabidopsis | SEFLPGFER. | ..VEDASSFP | LDYGIRRLDH | AVGNVP..EL | GPALTYVAGF | |
| Corn | EPFLPGFEG. | ..V..ASPGA | ADYGLSRFDH | IVGNVP..EL | APAAAYFAGF | |
| Rat | GRFLPGFEAP | TYKDTLLPKL | PSCNLEIIDH | IVGNQPDQEM | ESASEWYLKN | |
| Mouse | GRFLPGFEAP | TYKDTLLPKL | PRCNLEIIDH | IVGNQPDQEM | QSASEWYLKN | |
| Human | GQFLPGYEPP | AFMDPLLPLK | PKCSLEMIDH | IVGNQPDQEM | VSASEWYLKN | |
| Pig | GCFLPGFEAP | TFTDPLLSKL | PKCGLEIIDH | IVGNQPDQEM | ESASQWYMRN | |
| | | | | | | 300 |
| Arabidopsis | TGFHQFAEFT | ADDVGTAESG | LNSAVLASND | EMVLLPINEP | VHGTKRKSQI | |
| Corn | TGFHEFAEFT | TEDVGTAESG | LNSMVLANN | ENVLLPLNEP | VHGTKRKSQI | |
| Rat | LQFHREWSVD | DTQVHTEYSS | LRSIVVANYE | ESIKMPINEP | APG.RKKSQI | |
| Mouse | LQFHREWSVD | DTQVHTEYSS | LRSIVVTNYE | ESIKMPINEP | APG.RKKSQI | |
| Human | LQFHREWSVD | DTQVHTEYSS | LRSIVVANYE | ESIKMPINEP | APG.KKKSQI | |
| Pig | LQFHREWSVD | DTQIHTEYSA | LRSVVMANYE | ESIKMPINEP | APG.KKKSQI | |

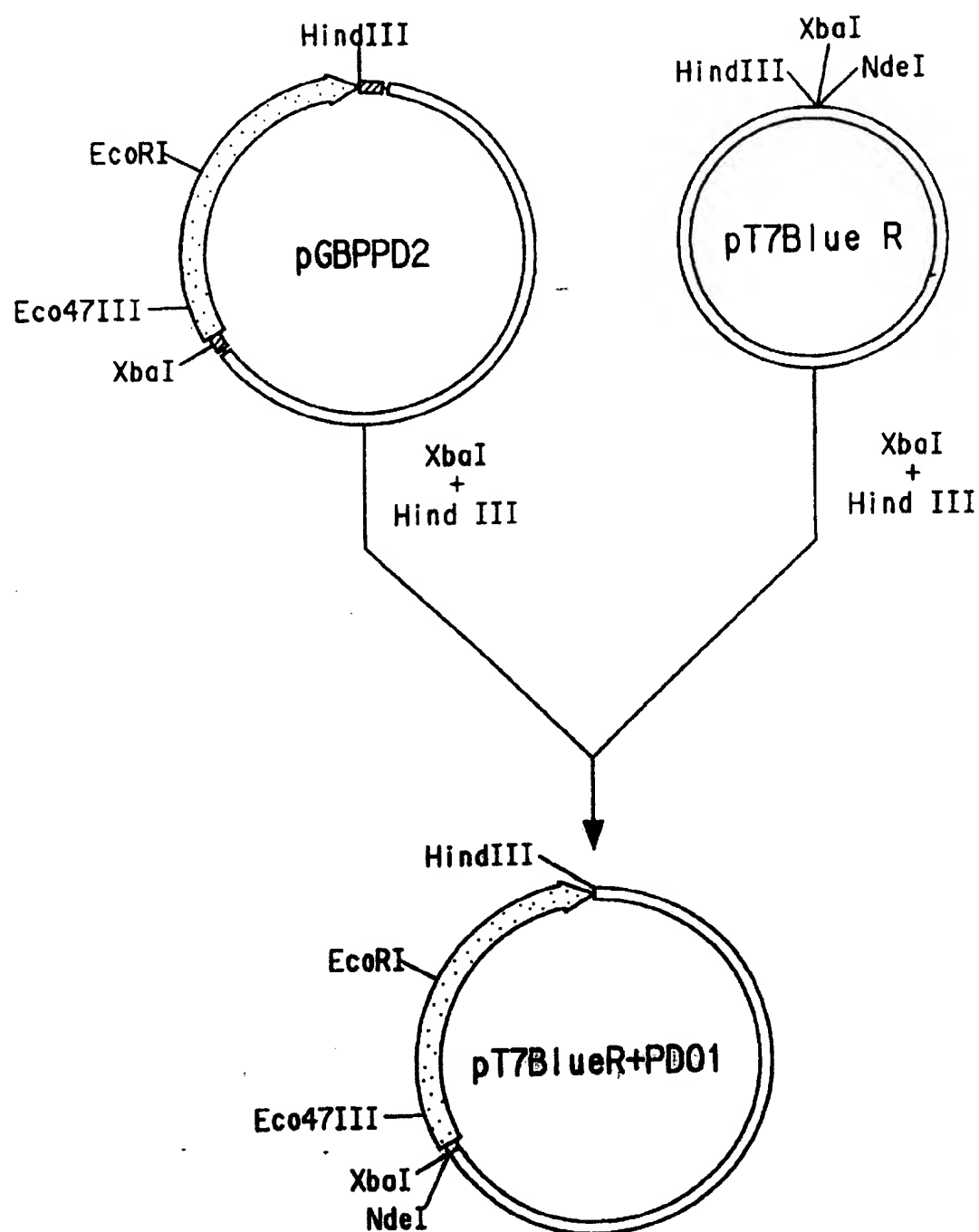
4/6

FIG. 3B

| | | | | |
|-------------|-------------|------------|------------|------------------------|
| | 301 | | | 350 |
| Arabidopsis | QTYLEHNEGA | GLQHLALMSE | DIFRTLREMR | KRSSIGGFDF NPSPPPTYQ |
| Corn | QTFLDHHGGP | GVQHMALASD | DVLRTLREMO | ARSAMGGFEF MAPPTSDYYD |
| Rat | QEYVDYNGGA | GVQHIALRTE | DIITTIRHLR | ER....GMEF LAMP.SSYYR |
| Mouse | QEYVDYNGGA | GVQHIALKTE | DIITAIRHLR | ER....GTEF LAAP.SSYYK |
| Human | QEYVDYNGGA | GVQHIALKTE | DIITAIRHLR | ER....GLEF LSVP.STYYK |
| Pig | QEYVDYNGGA | GVQHIALKTE | DIITAIRSLR | ER....GVEF LAMP.FTYYK |
| | . | . | . | . |
| | 351 | | | 400 |
| Arabidopsis | NLKK..RVGD | VLSDDQIKEC | EELGILVDRD | DOGTLLQIFT KPLGDRPTIF |
| Corn | GVRR..RAGD | VLTEAQIKEC | QELGVLVDRD | DOGVLLQIFT KPVGDRPTLF |
| Rat | LLRENLKTSK | IQVKENMDVL | EELKILVDYD | EKGYLELQIFT KPMQDRPTLF |
| Mouse | LLRENLKSAK | IQVKESMDVL | EELHILVDYD | EKGYLELQIFT KPMQDRPTLF |
| Human | QLREKLKTAK | IKVKENIDAL | EELKILVDYD | EKGYLELQIFT KPVQDRPTLF |
| Pig | QLQEKLSAK | IRVKESIDVL | EELKILVDYD | EKGYLELQIFT KPMQDRPTVF |
| | | | ** *** * | * * * * * ** * * * * |
| | 401 | | | 450 |
| Arabidopsis | IEIIQRVGCM | MKDEEGKAYQ | SGGCGGFGKG | NFSELFKSIE EYEKLEAKQ |
| Corn | LEIIQRIGCM | EKDEKQGEYQ | KGGCGGFGKG | NFSOLFKSIE DYEKSLEAKQ |
| Rat | LEVIQRHNNHQ | |GFGAG | NFNSLFKAFF E.EQALRG |
| Mouse | LEVIQRHNNHQ | |GFGAG | NFNSLFKAFF E.EQALRGNL |
| Human | LEVIQRHNNHQ | |GFGAG | NFNSLFKAFF E.EQNLRGNL |
| Pig | LEVIQRNNHQ | |GFGAG | NFNSLFKAFF E.EQELRGNL |
| | * *** | | *** | ** * * * * |
| | 451 | 462 | | |
| Arabidopsis | LVG | | | (Seq. I.D. No. 15) |
| Corn | AAAAAAAQGS | | | (Seq. I.D. No. 11) |
| Rat | | | | (Seq. I.D. No. 9) |
| Mouse | TDLEPNGVRS | GM | | (Seq. I.D. No. 8) |
| Human | TNMETNGVVP | GM | | (Seq. I.D. No. 6) |
| Pig | TDTDPNGVPP | RL | | (Seq. I.D. No. 7) |

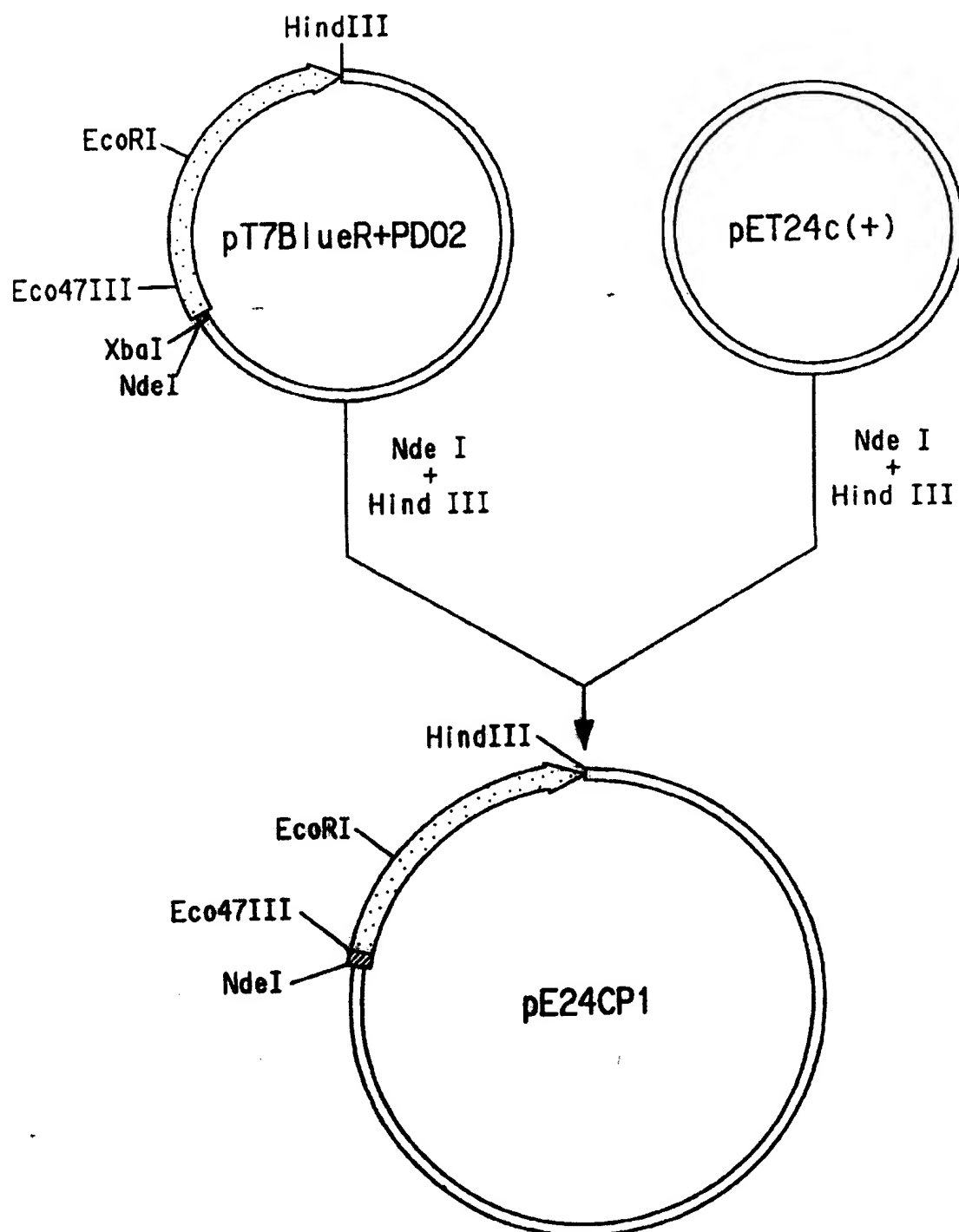
5/6

FIG. 4



6/6

FIG. 5



INTERNATIONAL SEARCH REPORT

Intern al Application No
PCT/US 97/11295

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/53 C12N15/82 C12Q1/26 C12Q1/02 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C12Q A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|--|-----------------------|
|------------|--|-----------------------|

| | | |
|---|--|-----|
| X | NEWMAN, T., ET AL.: "2960 Arabidopsis thaliana cDNA clone 91B13T7" EMBL SEQUENCE DATABASE, REL. 40, 16-JUN-1994, ACCESSION NO. T20952, XP002028637 see sequence | 1,2 |
| X | NEWMAN, T., ET AL.: "20804 Arabidopsis thaliana cDNA clone 231K20T7" EMBL SEQUENCE DATABASE, REL. 47, 8-MAR-1996, ACCESSION NO. N65764, XP002029449 see sequence | 1,2 |

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another claim or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents; such combination being obvious to a person skilled in the art.
- "Z" document member of the same patent family

Date of the actual completion of the international search

26 September 1997

Date of mailing of the international search report

07.10.97

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 apo nl
Fax: (+31-70) 340-3016

Authorized officer

Maddox, A

INTERNATIONAL SEARCH REPORT

Intern. Application No.

PCT/US 97/11295

| C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|--|--|-----------------------|
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| X | SCHULZ A ET AL: "SC-0051, A 2-BENZOYL-CYCLOHEXANE-1,3-DIONE BLEACHING HERBICIDE, IS APOTENT INHIBITOR OF THE ENZYME P-HYDROXYPHENYLPYRUVATE DIOXYGENASE" FEBS LETTERS, vol. 318, no. 2, March 1993, pages 162-166, XP002028049 see the whole document | 15 |
| X | --- BARTA I C ET AL: "BENZOYLCYCLOHEXANEDIONE HERBICIDES ARE STRONG INHIBITORS OF PURIFIED P-HYDROXYPHENYLPYRUVIC ACID DIOXYGENASE OF MAIZE" PESTICIDE SCIENCE, vol. 45, no. 3, 1 November 1995, page 286/287 XP000547268 see the whole document | 15 |
| X | --- EP 0 614 970 A (HOECHST SCHERING AGREVO GMBH) 14 September 1994 see the whole document | 15 |
| P,X | --- WO 96 38567 A (RHONE POULENC AGROCHIMIE ;SAILLAND ALAIN (FR); ROLLAND ANNE (FR);) 5 December 1996 see sequence ID no. 2 | 1,2 |
| P,X | --- BARTLEY, G.E., ET AL.: "Arabidopsis thaliana p-hydroxyphenylpyruvate dioxygenase (HPPD) cDNA, complete cds." EMBL SEQUENCE DATABASE, REL. 51, 19-MAR-1997, ACCESSION NO. U89267, XP002041908 see sequence | 1,2,20 |
| A | --- EP 0 652 286 A (RHONE POULENC AGROCHIMIE) 10 May 1995 see page 7, line 35 - line 47 | 10,16,18 |
| A | --- MISAWA N ET AL: "EXPRESSION OF AN ERWINA PHYTOENE DESATURASE GENE NOT ONLY CONFERS MULTIPLE RESISTANCE TO HERBICIDES INTERFERING WITH CAROTENOID BIOSYNTHESIS BUT ALSO ALTERS XANTHOPHYLL METABOLISM IN TRANSGENIC PLANTS" PLANT JOURNAL, vol. 6, no. 4, 1994, pages 481-489, XP002017203 see the whole document | 10,16,18 |
| | --- | |

-/--

INTERNATIONAL SEARCH REPORT

Intern: 21 Application No

PCT/US 97/11295

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|---|-----------------------|
| A | DENOYA C D ET AL: "A STREPTOMYCES AVERMITILIS GENE ENCODING A 4-HYDROXYPHENYLPYRUVIC ACID DIOXYGENASE-LIKE PROTEIN THAT DIRECTS THE PRODUCTION OF HOMOGENTISIC ACID AND AN OCHRONOTIC PIGMENT IN ESCHERICHIA COLI" JOURNAL OF BACTERIOLOGY, vol. 176, no. 17, September 1994, pages 5312-5319, XP002028042 see the whole document | 17 |
| A | NORRIS, S.R., ET AL.: "Gnetic dissection of carotenoid synthesis in Arabidopsis defines plastoquinone as an essential component of phytoene desaturation" THE PLANT CELL, vol. 7, December 1995, pages 2139-2149, XP002041909 cited in the application see the whole document | 1-20 |

INTERNATIONAL SEARCH REPORT

Information on patent family members

Intern. al Application No

PCT/US 97/11295

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---------------------|---|--|
| EP 0614970 A | 14-09-94 | DE 4305696 A CA 2116421 A JP 6343464 A | 01-09-94 26-08-94 20-12-94 |
| WO 9638567 A | 05-12-96 | FR 2734840 A FR 2734841 A FR 2734842 A AU 6228696 A | 06-12-96 06-12-96 06-12-96 18-12-96 |
| EP 0652286 A | 10-05-95 | FR 2712302 A AU 7775194 A BG 99169 A BR 9404562 A CA 2135461 A CN 1121958 A CZ 9402743 A HU 70464 A JP 7184664 A NZ 264879 A PL 305775 A SK 134094 A US 5635618 A ZA 9408826 A | 19-05-95 18-05-95 28-07-95 20-06-95 11-05-95 08-05-96 13-09-95 30-10-95 25-07-95 28-10-96 15-05-95 07-06-95 03-06-97 17-07-95 |